PENANGANAN PENCILAN DAN HETEROSKEDASTISITAS PADA MODEL REGRESI NONLINIER MENGGUNAKAN METODE *WEIGHTED LEAST SQUARE* DAN REGRESI *ROBUST*

(Studi Kasus Data Curah Hujan di Stasiun Meteorologi Maritim Tanjung Priok Tahun 2019-2020)

SKRIPSI

oleh:
RATIH KARTIKA RAHMATULNISSA
175090501111029

PROGRAM STUDI SARJANA STATISTIKA
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2021

PENANGANAN PENCILAN DAN HETEROSKEDASTISITAS PADA MODEL REGRESI NONLINIER MENGGUNAKAN METODE *WEIGHTED LEAST SQUARE* DAN REGRESI *ROBUST*

(Studi Kasus Data Curah Hujan di Stasiun Meteorologi Maritim Tanjung Priok Tahun 2019-2020)

SKRIPSI

oleh:
RATIH KARTIKA RAHMATULNISSA
175090501111029

PROGRAM STUDI SARJANA STATISTIKA
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS BRAWIJAYA
MALANG
2021

LEMBAR PENGESAHAN SKRIPSI
PENANGANAN PENCILAN DAN HETEROSKEDASTISITAS PADA MODEL REGRESI NONLINIER MENGGUNAKAN METODE *WEIGHTED LEAST SQUARE* DAN REGRESI *ROBUST*
(Studi Kasus Data Curah Hujan di Stasiun Meteorologi Maritim Tanjung Priok Tahun 2019-2020)

Oleh:
RATIH KARTIKA RAHMATULNISSA
175090501111029

Setelah dipertahankan di depan Majelis Penguji
pada tanggal 12 Juli 2021
dan dinyatakan memenuhi syarat untuk memperoleh gelar
Sarjana Statistika

Pembimbing,

Achmad Efendi, S.Si., M. Sc., Ph.D.
NIP. 198102192005011001

Mengetahui,
Ketua Jurusan Statistika
Fakultas MIPA Universitas Brawijaya

Rahma Fitriani, S.Si., M.Sc., Ph.D.
NIP. 197603281999032001

LEMBAR PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : RATIH KARTIKA RAHMATULNISSA
NIM : 175090501111029
PROGRAM STUDI : STATISTIKA
JUDUL SKRIPSI :

PENANGANAN PENCILAN DAN HETEROSKEDASTISITAS PADA MODEL REGRESI NONLINIER MENGGUNAKAN METODE *WEIGHTED LEAST SQUARE* DAN REGRESI *ROBUST*

(Studi Kasus Data Curah Hujan di Stasiun Meteorologi Maritim Tanjung Priok Tahun 2019-2020)

Dengan ini menyatakan bahwa:

1. **Isi dari Skripsi yang saya buat adalah benar-benar karya sendiri dan tidak menjiplak karya orang lain, selain nama-nama yang termasuk di isi dan tertulis di daftar pustaka dalam Skripsi ini.**
2. **Apabila dikemudian hari ternyata Skripsi yang saya tulis terbukti hasil jiplakan, amak saya akan bersedia menanggung segala resiko yang akan saya terima.**

Demikian pernyataan ini dibuat dengan segala kesadaran.

Malang, 12 Juli 2021
Yang Menyatakan,

Ratih Kartika Rahmatulnissa
NIM. 175090501111029

PENANGANAN PENCILAN DAN HETEROSKEDASTISITAS PADA MODEL REGRESI NONLINIER MENGGUNAKAN METODE *WEIGHTED LEAST SQUARE* DAN REGRESI *ROBUST*

(Studi Kasus Data Curah Hujan di Stasiun Meteorologi Maritim Tanjung Priok Tahun 2019-2020)

ABSTRAK

Analisis regresi merupakan metode yang dapat digunakan untuk menganalisis data untuk mengetahui hubungan antara dua atau lebih variabel independen terhadap variabel dependen. Analisis regresi berdasarkan linieritas dibagi menjadi dua yaitu regresi linier dan regresi nonlinier. Apabila hubungan variabel independen dan variabel dependen bersifat nonlinier maka data yang ditebarkan pada *scatterplot* tidak mengikuti garis lurus namun mengikuti suatu bentuk kurva tertentu. Selain itu, terdapat asumsi yang harus diuji yaitu asumsi normalitas galat, homoskedastisitas, multikolinieritas, dan autokorelasi. Pada penelitian ini, terdapat asumsi yang terlanggar yaitu asumsi homoskedastisitas dan pencilan. Oleh karena itu digunakan metode *Weighted Least Square* dan Regresi *Robust* untuk mengatasi terlanggarnya asumsi heteroskedastisitas dan pencilan. Penelitian ini menggunakan data sekunder yang didapatkan dari Badan Meterologi Klimatologi dan Geofisika mengenai curah hujan di DKI Jakarta Tahun 2019-2020 sebagai variabel dependen dan 4 variabel independen. Berdasarkan hasil penelitian ini dapat disimpulkan bahwa menggunakan metode *robust weighted least square* dapat menangani pencilan dan homoskedastisitas dimana nilai $R^2_{adjusted}$ sebesar 70,68% dan nilai *RMSE* sebesar 0,729.

Kata Kunci : Analisis regresi nonlinier, *Weighted Least Square*, Regresi Robust, LTS

HANDLING OF OUTLIER AND HETEROSCEDASTICITY ON NON LINEAR REGRESSION MODEL USING WEIGHTED LEAST SQUARE METHOD AND ROBUST REGRESSION

(Case Study of Rainfall Data in Stasiun Meteorologi Maritim Tanjung Priok 2019-2020)

ABSTRACT

Regression analysis is a method that can be used to determine the relationship between two or more independent variables on the dependent variable. Regression analysis based on linearity is divided into two, namely linear regression and nonlinear regression. If the relationship between the independent variable and the dependent variable is nonlinear, then the data spread on the scatterplot does not follow a straight line but follow a certain curve shape. In addition, there are assumptions that must be tested namely the assumptions of normality error, homoscedasticity, multicollinearity, and autocorrelation. In this study, there are assumptions that are violated, namely the assumption of homoscedastisity and outlier. Therefore, the Weighted Least Square and Robust Regression methods are used to overcome the heteroscedasticity and outlier. This study used secondary data were obtained from the Meteorology, Climatology, and Geophysics Agency regarding rainfall in DKI Jakarta in 2019-2020 as a dependent variable and 4 independent variables. Based on the results of this study it can be concluded that using Robust Weighted Least Square method can handle the outlier and heteroscedasticity where the adjusted R-Squared value is 70,68% and the RMSE value is 0,729.

Keywords: Nonlinear regression analysis, Weighted Least Square, Robust Regression, LTS

KATA PENGANTAR

Puji Syukur dipanjatkan kepada Tuhan Yang Maha Esa atas rahmat dan hidayah-Nya sehingga skripsi yang berjudul “Penanganan Pencilan Dan Heteroskedastisitas Pada Model Regresi Nonlinier Menggunakan Metode *Weighted Least Square* Dan Regresi *Robust*” sebagai salah satu syarat untuk memperoleh gelar Sarjana Statitsika dapat terselesaikan. Penyusunan skripsi ini tidak lepas dari berbagai bantuan, dukungan dan doa dari berbagai pihak. Oleh karena itu disampaikan banyak terima kasih kepada:

1. Bapak Achmad Efendi, S.Si., M. Sc., Ph.D. selaku dosen pembimbing skripsi yang telah memberikan bimbingan, bantuan, saran, dan motivasi selama proses penyusunan skripsi.
2. Ibu Dr. Ir. Maria Bernadetha Theresia M. selaku dosen penguji I yang telah memberikan bimbingan, arahan dan saran selama proses penyusunan skripsi.
3. Ibu Ir. Heni Kusdarwati, MS. selaku dosen penguji II yang telah memberikan bimbingan, arahan dan saran selama proses penyusunan skripsi.
4. Ibu Rahma Fitriani, S.Si., M. Sc., Ph.D. selaku Ketua Jurusan Statisitika Universitas Brawijaya.
5. Seluruh dosen, staf, karyawan Jurusan Statistika Fakultas MIPA Universitas Brawijaya.
6. Kedua orang tua saya, Almarhum bapak Margono dan ibu Indah, kakak saya Yoga Prassetya serta keluarga besar saya yang selalu memberikan dukungan dan doa.
7. Teman-teman seperbimbingan yang selalu memberikan dukungan dan semangat selama penyusunan skripsi.
8. Yoga, Audi, Melani dan Mita yang selalu mendengarkan suka dan duka serta memberikan semangat dalam pengerjaan skripsi.

9. Bella, Shelma, dan Ulfie yang menemani saya, membantu menyelesaikan kuliah, dan segala hal yang telah dilewati selama kuliah di Malang sejak maba hingga semester akhir.
10. Teman-seman Statistika 2017 yang senantiasa menemani dan membantu penulis sejak awal perkuliahan hingga semester akhir.

Penulis menyadari bahwa penyusunan skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis mengharapkan kritik dan saran demi perbaikan dan penyempurnaan dan semoga skripsi ini bermanfaat untuk bagi semua pihak.

Malang, 5 Mei 2021

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Analisis regresi adalah salah satu analisis statistika yang sering digunakan dalam penelitian. Menurut Drapper dan Smith (1992) analisis regresi merupakan metode yang dapat digunakan untuk menganalisis data dan mengambil kesimpulan yang bermakna mengenai hubungan ketergantungan variabel terhadap variabel lainnya. Analisis regresi yang menggunakan satu variabel independen merupakan analisis regresi linier sederhana dan yang menggunakan lebih dari satu variabel independen merupakan analisis regresi linier berganda.

Analisis regresi berdasarkan linieritas dibagi menjadi dua yaitu regresi linier dan regresi nonlinier. Garis regresi yang merupakan garis lengkung disebut dengan regresi nonlinier. Apabila hubungan antara variabel independen dengan variabel dependen bersifat nonlinier maka data yang ditebarkan pada *scatterplot* tidak mengikuti garis lurus namun mengikuti suatu bentuk kurva tertentu sehingga analisis regresi yang tepat untuk menerangkan hubungan variabel independen dan dependen adalah analisis regresi nonlinier.

Menurut (Chatterjee & Simonoff, 2013) Salah satu asumsi yang harus dipenuhi agar penduga dapat dipercaya adalah homoskedastisitas. Homoskedastisitas berarti varian dari sisaan bersifat konstan dan pelanggarannya disebut heteroskedastisitas. Salah satu metode yang digunakan untuk menyelesaikan kasus heteroskedastisitas adalah *Weighted Least Square.*

Menurut Montgomery dan Peck (1992) asumsi kenormalan seringkali tidak terpenuhi karena keberadaan pencilan yang memberikan pengaruh besar terhadap penduga parameter model. Penanganan masalah asumsi normalitas galat yang disebabkan adanya pencilan dapat ditangani menggunakaan regresi *robust* yang tahan terhadap pengaruh pencilan. Salah satu metode regresi *robust* dalam menangani pencilan ialah metode *Least Trimmed Square* yang mempunyai nilai *breakdown point* tinggi yaitu hampir 50%. Pada metode *LTS* pertama menyusun sisaan kuadrat dari nilai terkecil hingga terbesar, lalu mencari nilai h, dan banyak data yang menjadi estimasi *robust.*

Penelitian ini menggunakan data mengenai data curah hujan yang dipengaruhi oleh temperatur, kelembapan, lamanya penyinaran matahari, dan kecepatan angin di Stasiun Meteorologi Maritim Tanjung Priok periode bulan November 2019 hingga Februari 2020. Pada data BMKG seringkali yang merupakan hasil dari alam membentuk kurva lengkung sehingga pada penelitian ini menggunakan regresi nonlinier polinomial. Berdasarkan latar belakang yang telah dijabarkan, penelitian ini akan menangani pencilan dan heteroskedastisitas menggunakan metode *Weighted Least Square* dan Regresi *Robust LTS* pada regresi nonlinier yang diharapkan dapat memberikan solusi terbaik terhadap kedua kasus tersebut.

1.2 Rumusan Masalah

Rumusan Masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana penanganan asumsi heteroskedastisitas pada model regresi non linier ?
2. Bagaimana penanganan pencilan pada model regresi non linier ?

1.3 Tujuan

Tujuan yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Menduga parameter pada metode *Weighted Least Square* akibat data yang mengandung heteroskedastisitas pada model regresi non linier.
2. Menduga parameter pada model regresi non linier dengan menggunakan metode *Robust Least Trimmed Square*.

1.4 Manfaat

Manfaat yang diperoleh dari penelitian ini adalah sebagai berikut:

1. Mengetahui metode yang digunakan dalam data sehingga dapat digunakan pada analisis regresi nonlinier yang mengandung heteroskedastisitas dan pencilan.
2. Mengembangkan wawasan pembaca berkaitan dengan pemodelan regresi nonlinier yang melibatkan heteroskedastisitas dan pencilan.

1.5 Batasan Masalah

Batasan Masalah pada penelitian ini adalah sebagai berikut:

1. Permasalahan yang akan diselesaikan dalam penelitian ini adalah masalah heteroskedastisitas dan pencilan dalam analisis regresi nonlinier.
2. Data yang digunakan adalah data Stasiun Meterologi Kemayoran Tanjung Priok pada tahun 2019-2020.

BAB II
TINJAUAN PUSTAKA

2.1 Analisis Regresi

Menurut Gujarati (2003) istilah regresi diperkenalkan oleh Francis Galton. Analisis regresi berkaitan dengan studi tentang ketergantungan satu variabel dependen terhadap satu atau lebih variabel independen yang bertujuan untuk memprediksi nilai rata-rata populasi variabel dependen berdasarkan nilai tetap yang diketahui dari variabel independen.

Secara umum, analisis regresi dibagi menjadi dua yaitu regresi linier dan regresi nonlinier. Model regresi linier sederhana hanya melibatkan satu variabel independen dan menyatakan bahwa rata-rata sebenarnya dari variabel dependen berubah pada tingkat yang konstan ketika nilai variabel independen meningkat atau menurun. Persamaan model regresi linier sederhana disajikan pada persamaan (2.1).

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \tag{2.1}$$

Adapun model yang menggunakan lebih dari satu variabel independen untuk menjelaskan perilaku variabel terikat yaitu model regresi linier berganda yang disajikan pada persamaan (2.2).

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \varepsilon_i \tag{2.2}$$

di mana:

$i = 1,2,\ldots,n$

$j = 1,2,\ldots,k$

$n =$ banyaknya pengamatan

$k =$ banyaknya variabel independen

$Y_i =$ nilai pengamatan ke-i variabel dependen

$X_{ij} =$ nilai pengamatan ke-i variabel independen ke-j

$\beta_0 =$ intersep

$\beta_j =$ parameter regresi variabel independen ke-j

$\varepsilon_i =$ nilai ke-i sisaan di mana $\varepsilon_i \sim IIDN(0, \sigma^2)$

2.2 Regresi Nonlinier

Secara umum, model regresi dibagi menjadi dua yaitu model linier dan model nonlinier. Model regresi yang linier dapat dinyatakan pada persamaan (2.2). Apabila model tidak dapat dinyatakan dalam model tersebut maka model yang didapat adalah model nonlinier. Secara umum model nonlinier dapat dinyatakan dalam bentuk persamaan (2.3) sebagai berikut :

$$Y_i = f(x_i'; \theta) + \varepsilon_i, \tag{2.3}$$

Dimana:

$f(x_i'; \theta)$: fungsi regresi nonlinier dengan parameter θ yang harus diduga.

ε_i : nilai ke-i sisaan

i : amatan ke- i = 1,2,3...,n

Model nonlinier disajikan untuk mengilustrasikan fungsi yang menunjukkan bagaimana informasi pada sistem dapat digunakan untuk mengembangkan model yang lebih realistis (Rawling, Pantula dan Dickey, 1998).

Terdapat beberapa jenis model regresi nonlinier salah satunya adalah model polinomial yaitu model regresi nonlinier yang dibentuk dengan menjumlahkan pengaruh masing-masing variabel independen yang dipangkatkan secara meningkat hingga orde ke-k. Model regresi nonlinier dalam parameternya bersifat kuadratik dan kubik kurva yang dihasilkan membentuk garis lengkung. Regresi nonlinier model kuadratik merupakan hubungan antara dua variabel yang terdiri dari variabel dependen dan variabel independen yang akan diperoleh suatu kurva yang membentuk garis lengkung menaik $\beta_k > 0$ atau menurun $\beta_k < 0$. Bentuk persamaan model kuadratik secara umum pada persamaan (2.4) sebagai berikut (Steel dan Torrie,1980) :

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i, \tag{2.4}$$

Keterangan:

Y_i : nilai pengamatan ke-i variabel dependen

β_0 : intersep

β_1 dan β_2 : parameter regresi variabel independen

X_i : nilai pengamatan ke- i variabel independen

i : amatan ke- i = 1,2,3....,n

ε_i : nilai amatan ke- i sisaan,dimana $\varepsilon_i \sim IIDN(0, \sigma^2)$

Berikut persamaan model kuadratik dengan interaksi pada persamaan (2.5) sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_{11} {X_{1i}}^2 + \beta_{22} {X_{2i}}^2 + \beta_{12} X_{1i} X_{2i} + \varepsilon_i \quad (2.5)$$

Regresi polinomial terdapat dua persamaan yang umumnya digunakan yaitu orde 2 dan orde 3. Pemilihan persamaan regresi polinomial tergantung pada pola distribusi datanya.

2.3 Sifat-Sifat Penduga

Penduga parameter dikatakan penduga yang baik apabila memiliki sifat sebagai berikut:

a. Penduga yang tak bias

Penduga dikatakan tak bias apabila rata-rata penduga sama dengan nilai parameter.

$$E[\hat{\beta}] = \beta$$

b. Penduga yang efisien

Penduga dikatakan efisien apabila penduga tersebut memiliki ragam yang minimum. Apabila terdapat lebih dari satu penduga, peduga yang efisien adalah penduga yang memiliki ragam terkecil. Dua buah penduga dapat dibandingkan dengan efisiensi relatif. Misalkan $\hat{\beta}_1$ dan $\hat{\beta}_2$ adalah penduga untuk β, dimana ragam $\hat{\beta}_1$ lebih kecil dibanding ragam $\hat{\beta}_2$ maka $\hat{\beta}_1$ relatif lebih efisien dibandingkan dengan $\hat{\beta}_2$.

c. Penduga yang konsisten

Penduga dikatakan konsisten apabila nilai penduga $\hat{\beta}$ cenderung mendekati nilai parameter β untuk jumlah sampel yang semakin besar mendekati tak hingga. Jadi semakin besar ukuran sampel akan memberikan penduga parameter yang lebih baik dibandingkan ukuran sampel kecil.

2.4 Asumsi Linieritas

Asumsi linieritas digunakan untuk mengetahui apakah variabel independen dan variabel dependen memiliki hubungan yang linier secara signifikan. Salah satu metode dalam menguji asumsi linieritas adalah metide Ramsey RESET. Ramsey RESET menggunakan OLS untuk meminimumkan jumlah dari *error* yang dikuadratkan dari setiap amatan dengan langkah-langkah sebagai berikut (Gujarati, 2003):

a. Persamaan regresi pada persamaan (2.2) dapat dituliskan seperti berikut :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$

Lalu melakukan pendugaan parameter dengan pendekatan *Ordinary Least Square* (OLS) sehingga diperoleh penduga sebagai berikut :

$$\widehat{Y}_i = \widehat{\beta_0} + \widehat{\beta_1} X_{i1} + \widehat{\beta_2} X_{i2} + \cdots + \widehat{\beta_p} X_{ip}$$

Kemudian melakukan perhitungan R_1^2 pada persamaan (2.5) sebagai berikut :

$$R_1^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y}_i)^2} \quad (2.5)$$

b. Melakukan OLS untuk persamaan regresi kedua yaitu :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + \beta_{p+1} \widehat{Y}_i^{\,2} + \beta_{p+2} \widehat{Y}_i^{\,3} + \varepsilon_i \quad (2.6)$$

$$\widehat{Y}_i = \widehat{\beta_0} + \widehat{\beta_1} X_{i1} + \cdots + \widehat{\beta_p} X_{ip} + \beta_{p+1} \widehat{Y}_i^{\,2} + \beta_{p+2} \widehat{Y}_i^{\,3}$$

Kemudian menghitung perhitungan koefisien determinasi untuk menghasilkan R_2^2.

c. Melakukan pengujian bentuk hubungan variabel dependen dan variabel independen linier atau tidak linier yaitu :

Hipotesis:

$H_0 : \beta_{p+1} = \beta_{p+2} = 0$ *vs*

H_1 : minimal ada satu $a_j \neq 0$, j = p+1, p+2

Statistik Uji mengikuti sebaran F pada persamaan (2.7) sebagai berikut:

$$F = \frac{(R_2^2 - R_1^2)\ /\ jumlah\ variabel\ independen\ baru}{(1 - R_2^2)/(n - jumlah\ parameter\ di\ model\ baru)} \quad (2.7)$$

Keputusan untuk menolak H_0 apabila statistik uji $F > F_{(1-\alpha;2;(n-(p+2))}$ atau *p-value* < 0.05 yang berarti hubungan antara variabel independen dan variabel dependen adalah nonlinier.

2.5 Asumsi Klasik

2.5.1 Uji Normalitas Galat

Uji normalitas galat dilakukan untuk mengetahui residual dari data berdistribusi normal atau tidak. Apabila residual dari data tidak berdistribusi normal maka dapat menurunkan efisiensi penduga.

Uji normalitas galat dapat dilakukan dengan uji *Kolmogorov-Smirnov*. Menurut Siegel (1986) uji *Kolmogorov-Smirnov* didasarkan pada nilai deviasi maksimum. Berikut merupakan rumus uji *Kolmogorov-Smirnov* pada persamaan (2.8):

$$D = max|F_0(x_i) - S_n(x_i)|,\ i = 1,2,\ldots,n \tag{2.8}$$

dengan:

$F_0(x_i)$: Fungsi distribusi frekuensi kumulatif relatif dari distribusi teoritias dibawa H_0

$S_n(x_i)$: Distribusi frekuensi kumulatif pengamatan sebanyak sampel.

Hipotesis nol (H_0) pada uji *Kolmogorov-Smirnov* adalah sisaan berdistribusi normal. Kemudian kriteria untuk pengambilan keputusan adalah apabila $D < D_{tabel}$ maka dapat diputuskan untuk terima H_0 yang berarti bahwa asumsi normalitas terpenuhi.

Menurut Ghozali (2013), selain menggunakan uji *Kolmogorov-Smirnov* asumsi normalitas dapat dideteksi dengan melihat diagonal dari grafik residualnya dengan pengambilan keputusan sebagai berikut:

1. Jika data menyebar di sekitar garis diagonal maka pola menunjukkan berdistribusi normal.
2. Jika data menyebar jauh dari garis diagonal maka pola tidak menunjukkan berdistribusi normal.

2.5.2 Asumsi Homoskedastistias

Uji homoskedastisitas dilakukan untuk mengetahui galat memiliki ragam yang konstan. Pelanggaran dari asumsi ini berarti bahwa estimasi yang dilakukan tidak efisien dalam mengestimasi parameter yang sebenarnya (Chatterjee dan Simonoff, 2013). Salah satu uji homoskesdastisitas adalah dengan uji *Breusch-Pagan*. Uji tersebut didasarkan pada uji model regresi dan galat $u_{i,t}^2$ sebagai variabel dependen dan X sebagai variabel penjelas. Sehingga diperoleh *auxiliary regression* sebagai berikut:

$$u_{i,t}^2 = a_0 + a_1 X_{1i,t} + \cdots + a_k X_{ki,t} + v_i \tag{2.9}$$

Dimana :

$a_0, a_1, \ldots a_k$: koefisien dari *auxiliary regression*

v_i :galat dari *auxiliary regression*

Hipotesis pada uji *Breusch-Pagan* adalah:

H_0 : Var(ε_i) $= \sigma^2$ (ragam residual homogen) *vs*

H_1 : Var(ε_i) $\neq \sigma^2$ (ragam residual heterogen)

Statistik uji yang digunakan berdasarkan koefisien determinasi dari *auxiliary regression* adalah $LM = nR^2 \sim \chi_{k-1}^2$ dimana n adalah banyaknya amatan dan k banyak variabel independen. Kriteria pengambilan keputusan adalah terima H_0 apabila statistik uji > χ_{k-1}^2 sehingga dikatakan homoskedastisitas.

2.5.3 Asumsi Non-autokorelasi

Uji autokorelasi bertujuan untuk mengetahui apakah dalam model regresi terdapat korelasi tinggi antar $error_t$ dengan $error_{t-1}$.

$$Cov(\varepsilon_i, \varepsilon_j) = 0, i \neq j$$

$$\rho_{ij} = \frac{Cov(\varepsilon_i, \varepsilon_j)}{\sqrt{V(\varepsilon_i)V(\varepsilon_j)}} = \frac{0}{\sqrt{V(\varepsilon_i)V(\varepsilon_j)}} = 0$$

$$\rho_{ij} = 0$$

Pelanggaran terhadap asumsi kebebasan antar sisaan menyebabkan penduga ragam sisaan menjadi terlalu kecil dan tidak efisien. Statistik uji yang sering dipakai untuk mendeteksi autokorelasi dalam error adalah uji *Durbin Watson*.

Menurut Gujarati (2003) uji *durbin-watson* memiliki hipotesis sebagai berikut:

$H_0 : \rho_{ij} = 0$ (tidak terjadi autokorelasi antar sisaan) *vs*

$H_0 : \rho_{ij} \neq 0$ (terjadi autokorelasi antar sisaan)

Statistik pada *Durbin-Watson* dilambangkan dengan *d* dan hitung dengan persamaan (2.10)

$$d = \frac{\sum_{i=2}^{n}(\widehat{e_i} - \widehat{e_{i-1}})^2}{\sum_{i=1}^{n}\widehat{e_i}^2} \tag{2.10}$$

Statistik *d* tersebut dibandingan dengan nilai batas atas (d_U) dan nilai batas bawah (d_L) pada tabel durbin dengan α sebesar 5% dan 1% pada berbagai nilai n yang merupakan banyak amatan serta k yang merupakan banyak variabel independen.

Apabila nilai $d < d_L$ maka keputusan adalah menolak H_0 dengan kesimpulan terdapat autokorelasi. Apabila nilai yang diperoleh $d_L < d < d_U$ maka tidak dapat disimpulkan terjadi autokorelasi atau tidak sehingga diperlukan observasi lebih lanjut. Berikut kriteria pengambilan keputusan uji *Durbin-Watson.*

1. $d < d_L$, maka tolak H_0 (autokorelasi positif)
2. $d > 4 - d_L$, maka tolak H_0 (autokorelasi negatif)
3. $d_U < d < 4 - d_U$, maka terima H_0 (tidak terjadi autokorelasi)
4. $d_L \leq d \leq d_U$ atau $4 - d_U \leq d \leq 4 - d_L$ maka tidak dapat disimpulkan terjadi autokorelasi atau tidak.

2.5.4 Asumsi Non-Multikolinieritas

Multikolinieritas merupakan adanya hubungan linier antara variabel independen dalam suatu model regresi linier berganda. Menurut Gujarati (2003), permasalahan multikolinieritas menyebabkan nilai R^2 yang diperoleh tinggi namun banyak variabel independen yang tidak berpengaruh secara statistik.

Menurut Gujarati dan Porter (2009), untuk mendeteksi adanya multikolinieritas pada model regresi linier berganda dapat dilakukan dengan menggunakan nilai *Variance Inflation Factor (VIF).* Nilai *VIF* untuk koefisien regresi adalah persamaan (2.11) sebagai berikut:

$$VIF_i = \frac{1}{1-R_i^2} \quad ; \quad i = 1,2,\ldots,p \tag{2.11}$$

Di mana R_i^2 adalah nilai koefisien determinasi ketika *Xi* diregresikan dengan variabel independen lainnya. Apabila nilai *VIF* yang diperoleh lebih besar dari 10 (VIF > 10) maka diindikasikan terdapat multikolinieritas antar variabel independen.

Selain menggunakan *VIF,* untuk mendeteksi adanya multikolinieritas dapat melihat matriks korelasi dari variabel independen, jika koefisien korelasi > 0.8 maka terindikasi terdapat multikolinieritas (Gujarati, 2003).

2.6 *Weighted Least Square* (WLS)

2.6.1 *Weighted Least Square* (WLS)

Menurut Gujarati (2003) salah satu alternatif model estimasi untuk kasus heteroskedastisitas adalah metode *Weighted Least Square*. Hal ini dikarenakan metode WLS memiliki kemampuan untuk meminimumkan akibat dari pelanggaran asumsi.

Metode WLS merupakan kasus khusus dari GLS (*Generalized Least Square)*. Metode *Weighted Least Square* menggunakan *weight* atau pembobot yang proporsional terhadap kebalikan (*inverse)* dari ragam variabel dependen sehingga diperoleh error baru yang memiliki sifat pada regresi dengan OLS.

Metode ini ditujukan untuk meminimalkan jumlah dari kuadrat *error* setiap amatan sebagai berikut :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \varepsilon_i$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_K X_{ki})^2$$

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_K X_{ki})^2 \quad (2.12)$$

Nilai parameter β diperoleh dengan meminimumkan jumlah kuadrat residual yaitu dicari turunan dari $S(\beta)$ secara parsial terhadap β, dan disamakan dengan 0 dengan persamaan sebagai berikut:

$$\frac{\partial S}{\partial \beta_k} = -2 \sum_{i=0}^{n} X_{ki} \left(\sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \cdots - \beta_K X_{ki}) \right) = 0$$

Persamaan diatas menghasilkan persamaan normal sebagai berikut:

$$\sum_{i=1}^{n} X_{ki} Y_i = \widehat{\beta_0} \sum_{i=1}^{n} X_{ki} X_{0i} + \widehat{\beta_{k-1}} \sum_{i=1}^{n} X_{ki} X_{(k-1)i} + \widehat{\beta_k} \sum_{i=1}^{n} {X_{ki}}^2 = 0$$

$\sum_{i=1}^{n} X_{ki}Y_i - \widehat{\beta_0}\sum_{i=1}^{n} X_{ki}X_{0i} - \widehat{\beta_{k-1}}\sum_{i=1}^{n} X_{ki}X_{(k-1)i} - \widehat{\beta_k}\sum_{i=1}^{n} X_{ki}{}^2 = 0$

$\sum_{i=1}^{n} X_{ki}(Y_i - \widehat{\beta_0}X_{0i} - \cdots - \widehat{\beta_{k-1}}X_{(k-1)i} - \widehat{\beta_k}X_{ki}) = 0$

$$\sum_{i=1}^{n} X_{ji}\left[Y_i - \sum_{j=0}^{k} X_{ji}\hat{\beta}_j\right] = 0 \qquad (2.13)$$

Pendugaan parameter dengan metode WLS diperoleh dengan mengalikan fungsi pembobot W ke persamaan (2.14) diperoleh :

$\sum_{i=1}^{n} X_{ji}W_i\left[Y_i - \sum_{j=0}^{k} X_{ji}\hat{\beta}_j\right] = 0$

$\sum_{i=1}^{n} X_{ji}W_iY_i - \sum_{i=1}^{n}\sum_{j=0}^{k} X_{ji}W_iX_{ji}\hat{\beta}_j = 0$

$$\sum_{i=1}^{n}\sum_{j=0}^{k} X_{ji}W_iX_{ji}\hat{\beta}_j = \sum_{i=1}^{n} X_{ji}W_iY_i \qquad (2.14)$$

Apabila dituliskan dalam bentuk matriks penduga parameter WLS pada persamaan (2.15) sebagai berikut:

$\boldsymbol{X'WX\widehat{\beta}} = \boldsymbol{X'WY}$

$$\boldsymbol{\widehat{\beta}} = (\boldsymbol{X'WX})^{-1}\boldsymbol{X'WY} \qquad (2.15)$$

2.6.2 Pembobot bagi WLS

Penentuan pembobot dilakukan dengan melihat pola yang tunjukkan sisaan (*residual*) terhadap variabel independen. Pola tersebut antara lain (Gujarati, 2003):

- Varians *error* proporsional terhadap X_i^2

$$E(\varepsilon_i^2) = \sigma^2 X_i^2 \qquad (2.16)$$

Pembobot yang digunakan yaitu $\frac{1}{X_i}$ dengan persamaan regresi seperti berikut :

$$\frac{Y_i}{X_i} = \frac{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \varepsilon_i)}{X_i} \qquad (2.17)$$

- Varians *error* proporsional terhadap X_i

$$E(\varepsilon_i^2) = \sigma^2 X_i \qquad (2.18)$$

Pembobot yang digunakan yaitu $\frac{1}{\sqrt{X_i}}$ dengan persamaan regresi seperti berikut :

$$\frac{Y_i}{\sqrt{X_i}} = \frac{(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_j X_{ij} + \varepsilon_i)}{\sqrt{X_i}} \qquad (2.19)$$

Metode *Weighted Least Square* memiliki pembobot yang proporsional terhadap kebalikan (*inverse)* dari varian variabel dependen sehingga diperoleh error baru yang memiliki sifat pada regresi dengan metode *Ordinary Least Square*.

Pembobot dalam WLS menyebabkan penduga semakin kecil dibanding menggunakan OLS dan memperoleh ragam penduga semakin kecil. Ragam penduga yang semakin kecil menyebabkan penduga semakin baik dengan artian peduga berjarak sangat dekat dengan parameternya. Dengan begitu, dapat diperoleh ragam yang konstan sehingga dapat mengatasi masalah heteroskedastisitas pada model.

2.7 Identifikasi *Outlier*

Deteksi pencilan pada suatu analisis sangat diperlukan karena data yang memiliki karakteristik unik terlihat berbeda jauh dari amatan lainnya bisa jadi memiliki informasi yang penting dari sebuah data tersebut. Identifikasi *outlier* (pencilan) dapat dilakukan dengan perhitungan statistik sebagai berikut:

2.7.1 *Scatterplot*

Identifikasi *outlier* pada *scatterplot* dapat dilakukan dengan cara *plot residual* dengan nilai prediksi Y. Apabila terdapat data yang terletak jauh dari kumpulan data keseluruhan maka dapat mengindikasikan bahwa adanya pencilan dalam data tersebut. Namun hanya melihat *scatterplot* ini sangat rentan dalam penarikan kesimpulan karena penentuan pola pada plot bersifat subjektif.

Subjektif dalam artian ini bisa saja sebagian orang mengatakan plot tidak ada pola dan sebagian mengatakan ada pola. Maka dari itu, dalam identifikasi *outlier* yang menggunakan *scatterplot* keputusan hanya mengandalkan pengamatan atau penglihatan dari peneliti.

2.7.2 *Difference in Fitted Value* (DFFITS)

Menurut Kutner, dkk. (2004) identifikasi *outlier DFFITS* merupakan metode yang mendeteksi nilai pengamatan berpengaruh terhadap penduga Y. Nilai pengaruh h_{ii} dari nilai X didefinisikan sebagai elemen ke-*i* *hat matrix (H)* adalah:

$$h_{ii} = \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_i \qquad (2.20)$$

$\boldsymbol{x}_i = [1 \ X_{i1} \ X_{i2} \ ... \ X_{ij}]$ yang merupakan vektor baris berisi nilai variabel independen dalam amatan ke-*i* dan variabel independen ke-*k*. Maka nilai *DFFITS* dapat didefinisikan sebagai berikut :

$$DFFITS_i = e_i\sqrt{\frac{h_{ii}}{1-h_{ii}}}\ \sqrt{\frac{n-k-1}{JKG(1-h_{ii})-e_i^2}} \qquad (2.21)$$

dimana:

Dimana JKG merupakan jumlah kuadrat *residual* dan e_i merupakan residual amatan ke-i.

k : banyaknya parameter

n : banyaknya amatan.

Dengan Hipotesis nol (H_0) yang menyatakan pencilan ke-i tidak berpengaruh dan hipotesis alternatif (H_1) menyatakan pencilan ke-i berpengaruh maka kriteria pengambilan keputusan yang melandasi adalah sebagai berikut :

$$|DFFITS_i|\begin{cases} \leq 2\sqrt{\frac{p}{n}}, terima\ H_0 \\ > 2\sqrt{\frac{p}{n}}, tolak\ H_0 \end{cases}.$$

2.8 Regresi *Robust Least Trimmed Square*

2.8.1 Regresi Robust

Menurut Soemartini (2007) keberadaan pencilan menyebabkan residual yang besar dari model yang terbentuk, ragam pada data menjadi lebih besar, dan taksiran interval memiliki rentang yang lebar. Selain itu, pencilan berpengaruh memberikan nilai penduga yang bersifat bias sehingga interpretasi hasil yang diperoleh menjadi tidak valid.

Menurut Drapper dan Smith (1992) menjelaskan penyimpangan yang terjadi apabila terdapat pencilan adalah terlanggarnya asumsi normalitas. Membuang maupun menghapus suatu pencilan bukanlah prosedur yang bijaksana adakalanya pencilan dapat memberikan informasi yang tidak bisa diberikan oleh titik data lainnya. Untuk mengatasi penyimpangan asumsi normalitas galat serta kemungkinan kekurangan yang lain dapat digunakan regresi *robust* sebagai pengganti prosedur kuadrat terkecil.

Menurut Kutner, dkk. (2004) regresi *robust* dapat mengurangi pengaruh pencilan jika dibandingkan dengan MKT, sehingga penduga yang dihasilkan mempunyai sifat yang tidak berpengaruh terhadap keberadaan pencilan. Analisis regresi *robust* dapat digunakan untuk menganalisis dan mencocokkan model regresi dan mengatasi titik-titik pencilan yang memiliki nilai sisaan besar tanpa menghilangkan data tetapi model yang cocok dengan sebagian besar data sebagai solusi *robust*.

2.8.2 Regresi *Robust LTS*

Metode *Least Trimmed Square* (LTS) pertama kali diperkenalkan oleh Rousseeuw pada tahun 1984. Metode ini digunakan sebagai alternatif untuk mengatasi kelemahan metode OLS.

Metode LTS merupakan suatu metode pendugaan parameter pada regresi robust untuk meminimumkan jumlah kuadrat h residual. Persamaan metode ini sebagai berikut (Chen, 2002):

$$\widehat{\beta_{LTS}} = arg\ min \sum_i^h e_i^2 \qquad (2.22)$$

Dengan $h = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]$, $e_i = (\widehat{Y_i} - X_i\hat{\beta})$

Dimana :

e_i^2: kuadrat residual yang diurutkan dari terkecil ke terbesar ($e_1^2 < e_2^2, \dots, < e_n^2$)

n : banyak pengamatan

k : banyak variabel independen

Menurut Chen (2002) nilai h menunjukkan jumlah pengamatan yang digunakan untuk menduga parameter model regresi dan memberikan bobot nol pada *(n-h)* pengamatan. Secara tidak langsung *(n-h)* pengamatan dengan galat yang besar tidak akan mempengaruhi penduga parameter model. Nilai *h* yang optimal yang digunakan ialah $h = \left[\frac{n}{2}\right] + \left[\frac{k+2}{2}\right]$. Penggunaan h optimal akan mempengaruhi besar nilai *breakdown point. Breakdown point* adalah nilai yang menunjukkan proporsi terkecil dari data yang terpengaruh pencilan sehingga penduga bernilai jauh berbeda dengan penduga dari data yang tidak terpengaruh pencilan. Penduga LTS memiliki *breakdown point* yang tinggi yaitu sebesar 50%. Prosedur estimasi dengan estimasi LTS adalah sebagai berikut:

1. Mengestimasi koefisien regresi menggunakan MKT
2. Menentukan n residual $e_i{}^2 = (\hat{Y}_i - X_i\hat{\beta})^2$ kemudian menghitung jumlah $h = \left[\frac{n+k+2}{2}\right]$ pengamatan dengan nilai e_i^2 terkecil.
3. Menghitung $\sum_i^h e_i^2$
4. Mengestimasi parameter $\hat{\beta}_{baru}$ dari $\hat{\beta}_{observasi}$
5. Menentukan n kuadrat residual $e_i{}^2 = (\hat{Y}_i - X_i\hat{\beta})^2$ yang bersesuaian dengan $\hat{\beta}_{baru}$ kemudian menghitung sejumlah $\hat{\beta}_{baru}$ observasi dengan $e_i{}^2$ terkecil.
6. Menghitung $\sum_i^{h_{baru}} e_i^2$
7. Melakukan langkah dari tahap 4 sampai 6 untuk mendapatkan fungsi objektif yang kecil dan konvergen.

2.9 Uji Signifikansi Parameter

2.9.1 Uji Simultan

Menurut Ghozali (2013) Uji Simultan digunakan untuk mengetahui semua variabel independen yang dimasukkan dalam model memiliki pengaruh secara bersama-sama terhadap variabel dependen. Dengan tingkat signifikansi yang digunakan adalah alfa dengan derajat independen (*k-1,n-k-1*) maka hipotesis uji simultan sebagai berikut:

$H_0 : \hat{\beta}_1 = \hat{\beta}_2 = \cdots = \hat{\beta}_j = 0$ *vs*

$H_0 : \hat{\beta}_1 \neq \hat{\beta}_2 \neq \cdots \neq \hat{\beta}_j = 0$ (minimal satu β_j tidak sama dengan 0)

Statistik uji F :

$$fhit = \frac{R^2/k}{(1-R^2)/(n-k-1)} \tag{2.23}$$

Dengan kriteria keputusan sebagai berikut:

- $F_{hitung} < F_{tabel}$ (*P-value* > 0.05) maka menerima H_0 yang artinya variabel independen secara simultan tidak mempengaruhi variabel dependen secara signifikan.
- $F_{hitung} > F_{tabel}$ (*P-value* < 0.05) maka menolak H_0 yang artinya variabel independen secara simultan mempengaruhi variabel dependen secara signifikan.

2.9.2 Uji Parsial

Menurut Ghozali (2013) uji parsial dilihat dari nilai uji statistik t-student yang menunjukkan seberapa jauh pengaruh satu variabel independen secara individual dalam menerangkan variabel dependen. Dengan tingkat signifikansi 5% dan derajat independen (n-k-1) maka kriteria pengujian:

$H_0 : \beta_j = 0$ *vs*

$H_1 : \beta_j \neq 0$

Statistik uji :

$$thitung = \frac{\widehat{\beta_i} - \beta_j}{s_{\hat{\beta}_j}} \tag{2.24}$$

Dengan kriteria keputusan sebagai berikut:

- $t_{hitung} > t_{tabel}$ (*P-value* < 0.05) maka menolak H_0 yang artinya variabel independen secara parsial mempengaruhi variabel dependen secara signifikan.
- $t_{hitung} < t_{tabel}$ (*P-value* > 0.05) maka menerima H_0 yang artinya variabel independen secara parsial tidak mempengaruhi variabel dependen secara signifikan.

2.10 *Residual Standard Error (RSE)*

Keakuratan model dapat diukur dengan beberapa macam indikator. Salah satu ukuran tersebut adalah RSE yang merupakan simbangan baku (*standard deviation)* dari residual. Berikut merupakan rumus dari RSE:

$$RSE = \sqrt{\frac{SSE}{n-2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-p}} \quad (2.25)$$

Dimana:

$\hat{y}_i$= prediksi nilai y ke-i

n = jumlah sampel

p = jumlah parameter

SSE = jumlah kuadrat error

Suatu model dikatakan baik ketika nilai RSE yang diperoleh kecil atau mendekati nilai 0 (Montgomery, dkk., 2012).

2.11 *Mean Squared Error (MSE)*

Terdapat beberapa cara untuk melihat kebaikan model regresi salah satunya dengan melihat nilai MSE. Model yang baik adalah model regresi dengan nilai MSE yang kecil.

Berikut merupakan rumus MSE (James, dkk., 2013):

$$MSE = \frac{SSE}{n} \quad (2.26)$$

2.12 *Root Mean Squared Error (RMSE)*

Kebaikan model dapat dilihat menggunakan nilai RMSE. Model yang baik adalah model regresi dengan nilai RMSE yang kecil.

Berikut merupakan rumus RMSE :

$$RMSE = \sqrt{MSE} \quad (2.27)$$

2.13 Curah Hujan

Menurut Tanudidjaja (1993) Curah hujan adalah butir-butir air yang jatuh atau keluar dari awan atau kelompok awan. Apabila cuarahan dapat mencapai permukaan bumi disebut dengan hujan (Tjasyono,1999). Butir air yang keluar dari awan dan mencapai permukaan bumi harus memiliki garis tengah paling tidak sebesar 200 mikrometer. Apabila kurang dari ukuran diameter tersebut, butir-butir air yang dimaksud akan habis menguap di atmosfer sebelum mencapai permukaan bumi.

Banyaknya curah hujan yang mencapai permukaan bumi atau tanah selama selang waktu tertentu dapat diukur dengan jalan mengukur tinggi air hujan dengan cara tertentu. Hasil dari pengukuran tersebut dinamakan dengan curah hujan. Curah hujan merupakan salah satu unsur cuaca yang datanya diperoleh dengan cara mengukurnya dengan menggunakan alat penakar hujan, sehingga dapat diketahui jumlahnya dalam satuan millimeter (mm).

BAB III

METODOLOGI

3.1 Jenis dan Sumber Data

Dalam penelitian ini data yang digunakan adalah data sekunder yang diambil dari *website* BMKG. Data yang diperoleh yaitu data curah hujan di Stasiun Meteorologi Maritim Tanjung Priok pada bulan November 2019 hingga Februari 2020. Variabel yang digunakan pada penelitian ini adalah sebagai berikut:

Y : Curah hujan (mm)

$X1$: Temperatur rata-rata (ºC)

$X2$: Kelembapan rata-rata (%)

$X3$: Lamanya penyinaran matahari (jam)

$X4$: Kecepatan angin rata-rata (m/s)

3.2 Populasi dan Sampel

Populasi pada penelitian ini adalah data curah hujan seluruh stasiun BMKG di Indonesia tahun 2019-2020. Sampel yang digunakan yaitu data curah hujan Stasiun Meteorologi Maritim Tanjung Priok pada bulan November 2019 hingga Februari 2020.

3.3 Metode Analisis Data

Tahapan yang dilakukan pada penelitian ini sebagai berikut:

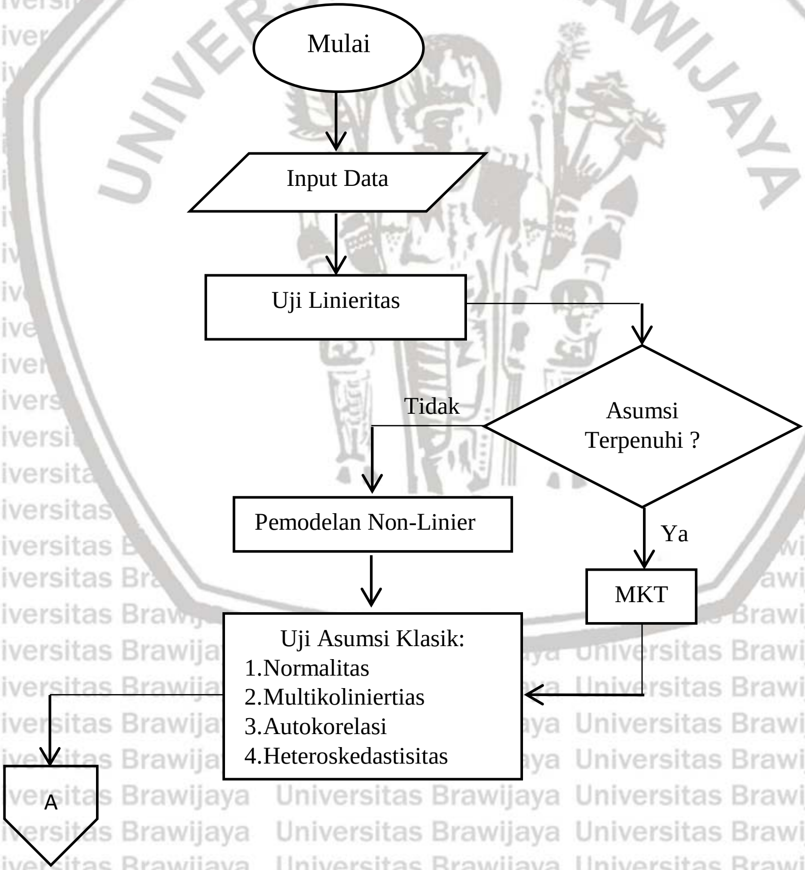
1. Mempersiapkan data yang akan digunakan sebagai variabel independen dan dependen.
2. Menguji linieritas data menggunakan *ramsey RESET test* untuk melihat hubungan setiap variabel independen terhadap variabel dependen sesuai rumus pada subbab 2.4.
3. Melakukan deteksi multikolinieritas untuk mendeteksi ada hubungan yang kuat antara variabel sesuai rumus pada subbab 2.5.4.
4. Melakukan pendugaan parameter dengan model non-linier polinomial sesuai persamaan 2.4.
5. Menguji asumsi klasik dengan data non-linier polinomial yaitu uji asumsi normalitas sisaan, non autokorelasi dan uji

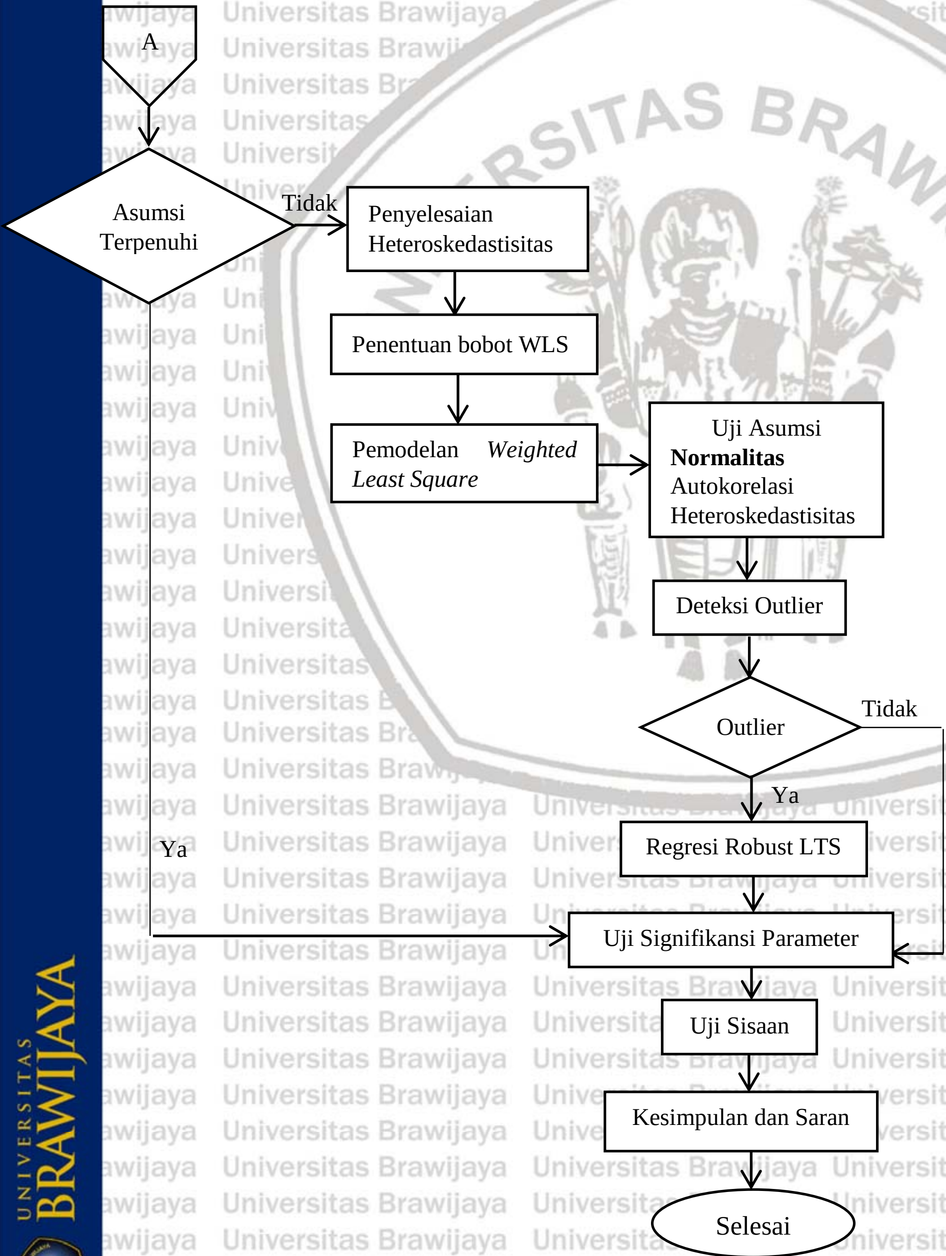
homoskedastisitas sesuai pada subbab 2.5. Apabila terdapat asumsi yang tidak terpenuhi yaitu asumsi homoskedastisitas diduga dengan metode *Weighted Least Squares* pada subbab 2.6.

6. Menangani asumsi homoskedastisitas yang tidak terpenuhi dapat dilakukan pembobot dengan metode penduga yang lain yakni *Weighted Least Squares* sesuai pada subbab 2.6.
7. Menentukan bobot *WLS* sesuai pada subbab 2.6.2.
8. Membentuk model *Weighted Least Squares* sesuai pada persamaan 2.15.
9. Melakukan pengujian asumsi normalitas sisaan. Apabila asumsi normalitas sisaan tidak terpenuhi dapat diduga adanya *outlier* dengan melanjutkan deteksi *outlier* sesuai pada subbab 2.7.
10. Mendeteksi *outlier* menggunakan identifikasi *DFFITS* sesuai pada subbab 2.7.2.
11. Menduga parameter regresi *robust LTS* sesuai pada subbab 2.8.
12. Menguji sisaan kembali untuk melihat asumsi normalitas galat, non-autokorelasi, homoskedastisitas.
13. Uji signifikansi parameter dengan uji simultan dan parsial sesuai pada subbab 2.9.
14. Menarik kesimpulan dan saran dari penduga *robust LTS*.

3.4 Diagram Alir

Diagram alir penelitian prosedur metode analisis data disajikan pada Gambar 3.1





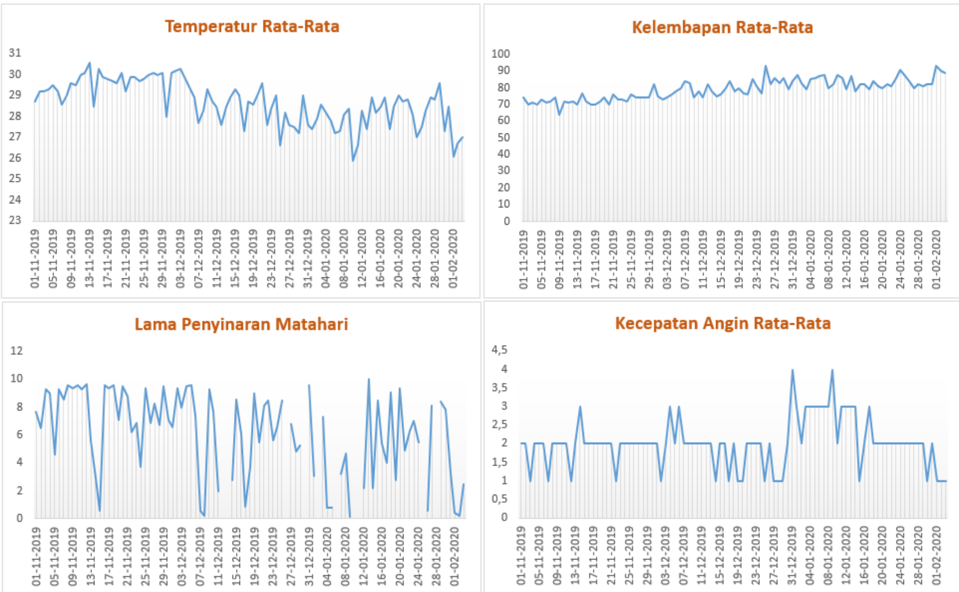
Gambar 3.1 Diagram Alir

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Deskriptif

Data yang digunakan dalam penelitian ini adalah data sekunder meliputi temperatur rata-rata, kelembapan rata-rata, lamanya penyinaran matahari, kecepatan angin rata-rata, dan curah hujan yang disajikan pada Lampiran 1. Analisis deskriptif digunakan untuk mengetahui gambaran umum mengenai data pada penelitian ini. Berikut merupakan plot data variabel independen yang disajikan pada Gambar 4.1 dan analisis deskriptif dari variabel penelitian yang disajikan pada Tabel 4.1.



Gambar 4.1 Plot Data Variabel Independen

Tabel 4.1 Analisis Deskriptif Variabel Penelitian

Variabel	Minimum	Maksimum	Rata-rata
Temperatur Rata-rata (ºC)	25,90	30,60	28,67
Kelembapan Rata-rata (%)	64	93	78,97
Lamanya penyinaran matahari (jam)	0,1	10	6,232
Kecepatan angin rata-rata (m/s)	1	4	2,011
Curah Hujan (mm)	0	146,1	18,6

Berdasarkan Tabel 4.1 dapat dilihat bahwa nilai rata-rata temperatur rata-rata di DKI Jakarta Stasiun Meterologi Maritim Tanjung Priok pada bulan November 2019 hingga Februari 2020 sebesar 28,67ºC. Dengan temperatur rata-rata paling rendah adalah 25,90ºC dan temperatur rata-rata paling tinggi adalah 30,60ºC. Nilai rata-rata kelembapan rata-rata sebesar 78,97% dengan tingkat kelembapan rendah sebesar 64% dan kelembapan tinggi sebesar 93%.

Dapat dilihat di DKI Jakarta Stasiun Meteorologi Maritim Tanjung Priok memiliki rata-rata lama penyinaran matahari sebesar 6,232 jam. Dapat diketahui pada Tabel 4.1 bahwa terdapat hari yang memiliki penyinaran paling sebentar yaitu sebesar 0,1 jam. Dan terdapat hari yang memiliki penyinaran matahari paling lama yaitu sebesar 10 jam.

Kecepatan angin rata-rata paling rendah adalah 1 m/s. Biasanya rata-rata kecepatan angin sebesar 2,011 m/s. Dan pada saat tertentu kecepatan angin rata-rata sampai 4 m/s. Dapat diketahui curah hujan paling sedikit adalah 0 mm yang artinya tidak terjadi hujan pada hari tersebut. Rata-rata curah hujan sebesar 18,60 mm dengan curah hujan tertinggi sebesar 146,1 mm.

4.2 Deteksi Multikolinieritas

Deteksi multikolinieritas pada penelitian ini bertujuan untuk mengetahui ada tidaknya korelasi antara variabel independen. Deteksi multikolinieritas dapat dilihat melalui *VIF* dan korelasi antar variabel. Pada penelitian ini, deteksi multikolinieritas menggunakan nilai korelasi antar variabel dan nilai *VIF*. Apabila nilai korelasi antar variabel > 0,9 maka terindikasi adanya multikolinieritas pada penelitian ini. Didapatkan nilai korelasi antar variabel disajikan pada Tabel 4.2 dan nilai *VIF* menggunakan *software R. 4.0.5.* pada Tabel 4.3 yang mengacu pada Lampiran 4.

Tabel 4.2 Nilai Korelasi antar Variabel Independen

Nilai Korelasi antar Variabel				
	X1	X2	X3	X4
X1	1	-0,8796764	0,4963874	-0,0132006
X2	-0,8796764	1	-0,4720335	0,0715837
X3	0,4963874	-0,4720335	1	-0,1719674
X4	-0,0132006	0,0715837	-0,1719674	1

Tabel 4.3 Nilai *VIF* Variabel Independen

Variabel	*VIF*
X1	4,684970
X2	4,511001
X3	1,382509
X4	1,051108

Berdasarkan Tabel 4.2 dapat diketahui bahwa nilai korelasi antar variabel independen tidak lebih dari 0,9 dan pada Tabel 4.3 nilai *VIF* tidak lebih dari 10 sehingga dapat dikatakan tidak terdapat masalah multikolinieritas dalam model.

4.3 Pengujian Asumsi Linieritas

Pengujian linieritas digunakan untuk melihat hubungan antar variabel independen terhadap variabel dependen. Pada penelitian ini digunakan Ramsey RESET *test* dengan kriteria apabila *p-value* lebih dari 0,05 maka hubungan antar variabel linier. Apabila hubungan antar variabel tidak linier maka dapat diuji menggunakan hubungan kuadratik sampai dengan kubik. Berikut merupakan hipotesis dalam uji linier.

H_0 : Variabel independen linier terhadap variabel dependen

H_1 : Variabel independen tidak linier terhadap variabel dependen

Hasil pengujian linieritas menggunakan program *R. 4.0.5* yang mengacu pada Lampiran 3 disajikan pada Tabel 4.4.

Tabel 4.4 Uji Linieritas

Variabel	*P-value*			Kesimpulan
	Linier	Kuadratik	Kubik	
Temperatur rata-rata	0,404	0,591	0,180	Linier
Kelembapan rata-rata	0,000	0,983	0,984	Kuadratik
Lama penyinaran matahari	0,917	0,758	0,898	Linier
Kecepatan angin rata-rata	0,033	0,966	1	Kuadratik

Berdasarkan Tabel 4.4 dapat ditentukan bahwa hubungan antara variabel independen terhadap variabel dependen secara berturut-turut adalah linier, kuadratik, linier, dan kuadratik. Dari hasil tersebut dapat digunakan untuk pembentukan model regresi non-linier dan pendugaan parameter.

4.4 Pembentukan Model Regresi Non-Linier menggunakan MKT

Pada subbab 4.2 data pada penelitian ini tidak terjadi masalah multikolinieritas sehingga dapat dilanjutkan dengan pendugaan parameter menggunakan MKT. Dengan menggunakan program *R. 4.0.5* yang mengacu pada Lampiran 5 didapatkan statistik pada Tabel 4.5.

Tabel 4.5 Statistik dan Uji Signifikansi Parameter Model Regresi Non-linier

Parameter	Koefisien	*Standar error*	Nilai t	*p-value*
β_0 (Intersep)	1408,250	339,25910	4,151	7,63e-05 ***
β_1 (X_1)	-4,489	4,83025	-0,929	0,355216
β_2 (X_2)	-32,829	8,64752	-3,796	0,000269 ***
β_3 $({X_2}^2)$	0,214	0,05559	3,849	0,000225 ***
β_4 (X_3)	0,585	0,82098	0,585	0,478056
β_5 (X_4)	-18,265	15,24291	-1,198	0,234039
β_6 $({X_4}^2)$	4,299	3,44420	1,248	0,215240

Berdasarkan tabel 4.4 didapatkan pemodelan menggunakan MKT sebagai berikut:

$$y = 1408{,}25 - 4{,}49x_1 - 32{,}83x_2 + 0{,}21395x_2^2 + 0{,}585x_3 - 18{,}265x_4 + 4{,}299x_4^2$$

Pada program R 4.0.5 didapatkan R^2 sebesar 32,01% keragaman curah hujan dipengaruhi oleh temperatur rata-rata, kelembapan rata-rata, lamanya penyinaran matahari, dan kecepatan angin rata-rata sedangkan 67,99% dijelaskan oleh variabel lain diluar analisis ini.

4.4.1 Pengujian Homoskedastisitas

Uji homoskedastisitas digunakan untuk mengetahui galat memiliki ragam yang konstan. Pengujian homoskedastisitas dapat dilihat menggunakan uji *Breusch-Pagan* dengan hipotesis:

Hipotesis
H_0 : $\text{Var}(\varepsilon_i) = \sigma^2$ (ragam residual homogen) *vs*
H_1 : $\text{Var}(\varepsilon_i) \neq \sigma^2$ (ragam residual heterogen)

Berdasarkan hasil pengujian menggunakan program *R 4.0.5* yang mengacu pada Lampiran 5 memberikan hasil nilai *Breusch-Pagan* sebesar 17,515 dengan *p-value* sebesar 0,007565 < 0,05.

Berdasarkan *p-value* dapat disimpulkan bahwa keputusan menolak H_0 yang berarti ragam galat tidak homogen atau terjadi masalah heteroskedastisitas pada model regresi.

4.4.2 Pengujian Normalitas

Uji normalitas digunakan untuk mengetahui residual dari data berdistribusi normal atau tidak. Pada penelitian ini, digunakan uji *Kolmogorov-Smirnov* pada subbab 2.5.1. Berikut hipotesis yang digunakan pada uji normalitas:
H_0 : Galat berdistribusi normal *vs*
H_1 : Galat tidak berdistribusi normal

Berdasarkan hasil pengujian normalitas galat menggunakan program *R. 4.0.5* yang mengacu pada Lampiran 6 memberikan hasil nilai *Kolmogorov-smirnov* sebesar 0,21956 dengan *p-value* sebesar 2,934e-12 < 0,05.

Berdasarkan *p-value* < 0,05 dapat disimpulkan keputusan menolak H_0 yang berarti galat tidak berdistribusi normal.

4.4.3 Pengujian Autokorelasi

Uji Autokorelasi bertujuan untuk mengetahui apakah dalam model regresi terdapat korelasi tinggi antar error ke t dengan error ke t-1. Pengujian autokorelasi dapat menggunakan uji *Durbin Watson* dengan hipotesis sebagai berikut:
H_0 : $\rho_{ij} = 0$ (tidak terjadi autokorelasi antar sisaan) *vs*
H_0 : $\rho_{ij} \neq 0$ (terjadi autokorelasi antar sisaan)

Tabel *durbin-watson* dengan $k = 4$ dan $n = 95$ didapatkan nilai $DL = 1{,}5795$ dan $DU = 1{,}7546$. Hasil pengujian autokorelasi menggunakan uji Durbin Watson tersaji pada Tabel 4.6.

Tabel 4.6 Hasil Uji Autokorelasi

Nilai Uji Asumsi	
DL	1,5795
d	2.4417
DU	1,7546
P-value	0,9790

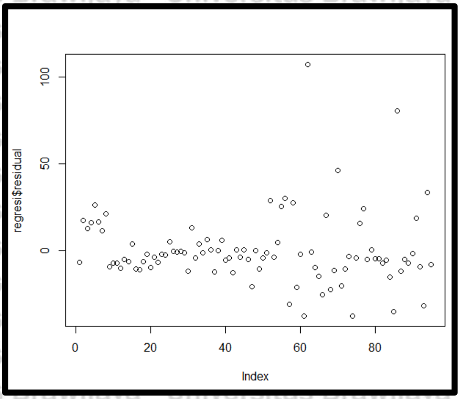
Selanjutnya pada Tabel 4.6 didapatkan nilai durbin watson sebesar 2,4417 yang dibandingkan dengan tabel *durbin-watson.* Nilai $d > DU$ sehingga dapat disimpulkan tidak terjadi autokorelasi positif pada data penelitian ini.

4.5 Deteksi Outlier

Setelah pengujian normalitas pada subbab 4.4.2 yang memberikan kesimpulan galat tidak berdistribusi normal dapat dilanjutkan dengan deteksi *outlier*. Penyebab galat tidak berdistribusi normal dapat disebabkan adanya *outlier* yang tidak menyebar secara distribusi normal. Oleh karena itu dilakukan deteksi *outlier.*

4.5.1 *Scatterplot*

Plot residual data curah hujan di DKI Jakarta Stasiun Meteorologi Maritim Tanjung Priok tersaji pada Gambar 4.1.



Gambar 4.1 *Scatterplot* Residual Curah Hujan

Pada Gambar 4.1 dapat dilihat bahwa terdapat titik-titik residual yang tidak berada di sebaran titik lainnya. Titik tersebut disebut dengan *outlier* karena menyebar jauh dari sebaran data yang ada.

4.5.2 *DFFITS*

DFFITS merupakan deteksi *outlier* yang mendeteksi nilai amatan berpengaruh terhadap penduga Y dengan hipotesis sebagai berikut :

H_0 : Pencilan ke-i tidak berpengaruh *vs*

H_1 : Pencilan ke-i berpengaruh

Dengan program *R 4.0.5* yang mengacu pada Lampiran 8 didapatkan nilai *DFFITS* pada Tabel 4.7.

Tabel 4.7 Nilai DFFITS

i	*DFFITS*	Nilai *Cut-off*	Keputusan
55	0,8134949711	0,410391	Tolak H_0 (pencilan ke-i berpengaruh)
61	-1,7685683636		
62	1,7480911841		
70	2,0782990608		
71	-0,4350847569		
74	-0,7147640997		
85	-0,7393179199		
86	0,9382369780		
93	-0,9111023862		
94	0,6334564370		

Nilai $DFFITS_i$ dibandingkan dengan nilai *cutoff* yaitu $2\sqrt{\frac{k}{n}}$ dengan $k = 4$ dan $n = 95$ didapatkan nilai *cutoff* sebesar 0,410391. Setelah itu, nilai $DFFITS_i$ yang didapatkan pada Tabel 4.8

dibandingkan dengan nilai *cuttoff*. Apabila $DFFITS_i > 2\sqrt{\frac{k}{n}}$ maka pencilan ke-i berpengaruh. Berdasarkan Tabel 4.8 dapat disimpulkan data yang merupakan pencilan yang berpengaruh adalah data ke-55,70,71,74,85,86,93,dan ke-94.

4.6 *Weighted Least Square*

Berdasarkan subbab 4.5 asumsi homoskedastisitas terlanggar yang berarti ragam dari sisaan tidak bersifat homogen. Pelanggaran asumsi ini dapat mengakibatkan penduga yang diperoleh varian cenderung membesar. Oleh karena itu pada penelitian ini dilakukan *weighted least square* untuk menyelesaikan kasus heteroskedastisitas.

Pada penelitian ini menggunakan 8 jenis pembobot yaitu sebagai berikut $1/_{X1}$, $1/_{X2}$, $1/_{X3}$, $1/_{X4}$, $1/_{\sqrt{X1}}$, $1/_{\sqrt{X2}}$, $1/_{\sqrt{X3}}$, $1/_{\sqrt{X4}}$. Penentuan bobot disesuaikan dengan pola residual yaitu ragam residual proporsional terhadap Xi^2 dan ragam residual proporsional terhadap Xi sehingga bobot dalam metode *Weighted Least Square* pada penelitian ini adalah $1/_{Xi}$ dan $1/_{\sqrt{Xi}}$.

Setelah memberikan pembobot terhadap persamaan yang digunakan, dilakukan uji homoskedastisitas *breusch-pagan* yang tersaji pada Tabel 4.8.

Tabel 4.8 Pengujian Homoskedastisitas

Persamaan	Nilai *Breusch Pagan*	*P-value*	Keputusan
1	18,243	0,05099	Homoskedastisitas
3	55,148	2,963e-08	Heteroskedastisitas
5	17,911	0,05649	Homoskedastisitas
7	35,337	0,00011	Heteroskedastisitas

Berdasarkan Tabel 4.10 dapat diketahui persamaan 1 dan persamaan 5 memiliki ragam galat yang konstan (homoskedastisitas).

Penduga parameter persamaan 1 disajikan pada tabel 4.9 dan penduga parameter persamaan 5 disajikan pada tabel 4.10

Tabel 4.9 Statistik dan Signifikansi Parameter Model *Weighted Least Square* Persamaan 1

Parameter	Koefisien	*P-value*
β_0 (Intersep)	62,290	9,7e-05 ***
β_1 (X_1)	-0,256	0,154389
β_2 (X_2)	-1,362	0,000181 ***
β_3 $({X_2}^2)$	0,009	0,000111 ***
β_4 (X_3)	-0,075	0,663139
β_5 (X_4)	-1,645	0,113854
β_6 $({X_4}^2)$	0,462	0,039782 *
β_7 $(X_1 \times X_2 \times X_3 \times X_4)$	-8,369e-05	0,737732
β_8 $(X_1 \times {X_2}^2 \times X_3 \times X_4)$	2,101e-06	0,463904
β_9 $(X_1 \times X_2 \times X_3 \times {X_4}^2)$	2,099e-05	0,821098
β_{10} $(X_1 \times {X_2}^2 \times X_3 \times {X_4}^2)$	-5,926e-07	0,584503

Tabel 4.10 Statistik dan Signifikansi Parameter Model *Weighted Least Square* Persamaan 5

Parameter	Koefisien	*P-value*
β_0 (Intersep)	330,200	0,000137 ***
β_1 (X_1)	-1,410	0,147448
β_2 (X_2)	-7,170	0,000224 ***
β_3 $({X_2}^2)$	0,045	0,000138 ***
β_4 (X_3)	-0,340	0,704475
β_5 (X_4)	-8,387	0,122705

Parameter	Koefisien	*P-value*
$\beta_6\ (X_4{}^2)$	2,386	0,042519 *
$\beta_7\ (X_1 \times X_2 \times X_3 \times X_4 \times \sqrt{X}_1)$	-1,094e-04	0,649804
$\beta_8\ (X_1 \times X_2{}^2 \times X_3 \times X_4 \times \sqrt{X}_1)$	2,364e-06	0,401064
$\beta_9\ (X_1 \times X_2 \times X_3 \times X_4{}^2 \times \sqrt{X}_1)$	2,836e-05	0,753606
$\beta_{10}\ (X_1 \times X_2{}^2 \times X_3 \times X_4{}^2 \times \sqrt{X}_1)$	-6,726e-07	0,526202

Berdasarkan hasil dari program *R. 4.0.5* yang mengacu pada Lampiran 9 persamaan 1 memiliki nilai $R^2_{adjusted}$ sebesar 38,52% sedangkan pada persamaan 5 memiliki nilai $R^2_{adjusted}$ sebesar 37,79% sehingga melalui nilai $R^2_{adjusted}$ dapat disimpulkan bahwa persamaan 1 memiliki nilai lebih besar dibanding persamaan 5. Setelah itu dilakukan pengecekan asumsi normalitas pada persamaan 1 dan persamaan 5.

4.6.1 Pengecekan Asumsi Normalitas

Uji normalitas digunakan untuk mengetahui residual dari data berdistribusi normal atau tidak. Pada penelitian ini, digunakan uji *Kolmogorov-Smirnov* pada subbab 2.5.1. Berikut hipotesis yang digunakan pada uji normalitas:

H_0 : Galat berdistribusi normal *vs*

H_1 : Galat tidak berdistribusi normal

Berdasarkan hasil pengujian normalitas galat menggunakan program *R. 4.0.5* yang mengacu pada Lampiran 10 pada persamaan 1 memiliki nilai *Kolmogorov-smirnov* sebesar 0,14166 dengan *p-value* sebesar 6,993e-05 < 0,05. Karena *p-value* < 0,05 dapat disimpulkan keputusan menolak H_0 yang berarti galat tidak berdistribusi normal.

Berdasarkan hasil pengujian normalitas galat menggunakan program *R. 4.0.5* yang mengacu pada Lampiran 10 pada persamaan 5 memiliki nilai *Kolmogorov-smirnov* sebesar 0,14199 dengan *p-value* sebesar 6,627e-05 < 0,05. Karena *p-value* < 0,05 dapat disimpulkan keputusan menolak H_0 yang berarti galat tidak berdistribusi normal sehingga dilanjutkan dengan penanganan pencilan menggunakan regresi robust dengan metode *LTS (Least Trimmed Square).*

4.7 Regresi *Robust*

Regresi *robust* berkaitan dengan penyelesaian masalah *outlier* menggunakan metode LTS. Berikut merupakan penduga parameter persamaan 1 yang disajikan pada tabel 4.11 dan penduga parameter persamaan 5 yang disajikan pada tabel 4.12.

Tabel 4.11 Statistik dan Signifikansi Parameter Regresi *Robust WLS* Persamaan 1

Parameter	Koefisien	*P-value*
β_0 (Intersep)	-3,8570	0,474
β_1 (X_1)	0,0567	0,172
β_2 (X_2)	0,0437	0,736
β_3 $({X_2}^2)$	-1,668e-05	0,838
β_4 (X_3)	-0,0070	0,886
β_5 (X_4)	-0,0487	0,882
β_6 $({X_4}^2)$	0,0033	0,966
β_7 $(X_1 \times X_2 \times X_3 \times X_4)$	-2,570e-05	0,721
β_8 $(X_1 \times {X_2}^2 \times X_3 \times X_4)$	3,319e-07	0,667
β_9 $(X_1 \times X_2 \times X_3 \times {X_4}^2)$	1,331e-05	0,585
β_{10} $(X_1 \times {X_2}^2 \times X_3 \times {X_4}^2)$	-1,646e-07	0,550

Tabel 4.12 Statistik dan Signifikansi Parameter Regresi *Robust WLS* Persamaan 5

Parameter	Koefisien	*P-value*
β_0 (Intersep)	92,730	0,000717 ***
β_1 (X_1)	-0,351	0,172461
β_2 (X_2)	-2,028	0,001162 **
β_3 $({X_2}^2)$	0,014	0,000387 ***
β_4 (X_3)	-1,131	0,000872 ***

Parameter	Koefisien	*P-value*
$\beta_5\ (X_4)$	-6,618	0,002005 **
$\beta_6\ (X_4{}^2)$	1,495	0,002265 **
$\beta_7\ (X_1 \times X_2 \times X_3 \times X_4 \times \sqrt{X}_1)$	2,161e-04	0,007133 **
$\beta_8\ (X_1 \times X_2{}^2 \times X_3 \times X_4 \times \sqrt{X}_1)$	-1,553e-06	0,083502
$\beta_9\ (X_1 \times X_2 \times X_3 \times X_4{}^2 \times \sqrt{X}_1)$	-6,108e-05	0,029771 *
$\beta_{10}\ (X_1 \times X_2{}^2 \times X_3 \times X_4{}^2 \times \sqrt{X}_1)$	4,777e-07	0,146732

Berdasarkan Tabel 4.11 yang diolah menggunakan *software R. 4.0.5* yang mengacu pada Lampiran 11 didapatkan nilai R^2 sebesar 16,14% dan $R^2_{adjusted}$ sebesar 1,43%. Sedangkan pada Tabel 4.12 didapatkan nilai R^2 sebesar 74,99% dan $R^2_{adjusted}$ sebesar 70,68%. Nilai $R^2_{adjusted}$ regresi *robust WLS* pada tabel 4.12 sudah meningkat dibandingkan nilai $R^2_{adjusted}$ pada model WLS maupun MKT. berdasarkan Tabel 4.12 didapatkan variabel yang berpengaruh terhadap curah hujan adalah kelembapan rata-rata, lama penyinaran matahari dan kecepatan angin rata-rata.

4.7.1 Asumsi Normalitas

Uji normalitas digunakan untuk mengetahui residual dari data berdistribusi normal atau tidak. Pada penelitian ini, digunakan uji *Kolmogorov-Smirnov* pada subbab 2.5.1. Berikut hipotesis yang digunakan pada uji normalitas:

H_0 : Galat berdistribusi normal *vs*

H_1 : Galat tidak berdistribusi normal

Berdasarkan hasil pengujian normalitas galat menggunakan program *R. 4.0.5* yang mengacu pada Lampiran 12 memberikan hasil nilai *Kolmogorov-smirnov* sebesar 0,27608 dengan *p-value* sebesar 2,2e-16 < 0,05.

Berdasarkan *p-value* dapat disimpulkan bahwa keputusan menerima H_0 yang berarti ragam tidak berdistribusi normal pada model regresi.

4.7.2 Asumsi Heteroskedastisitas

Uji homoskedastisitas digunakan untuk mengetahui galat memiliki ragam yang konstan. Pengujian homoskedastisitas dapat dilihat menggunakan uji *Breusch-Pagan* dengan hipotesis:

H_0 : $\text{Var}(\varepsilon_i) = \sigma^2$ (ragam residual homogen) *vs*

H_1 : $\text{Var}(\varepsilon_i) \neq \sigma^2$ (ragam residual heterogen)

Berdasarkan hasil pengujian menggunakan program *R 4.0.5* yang mengacu pada Lampiran 12 memberikan hasil nilai *Breusch-Pagan* sebesar 17,911 dengan *p-value* sebesar 0,05649 > 0,05.

Berdasarkan *p-value* dapat disimpulkan bahwa keputusan menerima H_0 yang berarti ragam galat homogen atau tidak terjadi masalah heteroskedastisitas pada model regresi.

4.7.3 Asumsi Autokorelasi

Uji Autokorelasi bertujuan untuk mengetahui apakah dalam model regresi terdapat korelasi tinggi antar *error* ke t dengan *error* ke t-1. Pengujian autokorelasi dapat menggunakan uji Durbin Watson dengan hipotesis sebagai berikut:

H_0 : $\rho_{ij} = 0$ (tidak terjadi autokorelasi antar sisaan) *vs*

H_0 : $\rho_{ij} \neq 0$ (terjadi autokorelasi antar sisaan)

Tabel *durbin-watson* dengan $k = 4$ dan $n = 95$ didapatkan nilai $DL = 1{,}5795$ dan $DU = 1{,}7546$. Hasil pengujian autokorelasi menggunakan uji Durbin Watson tersaji pada Tabel 4.13.

Tabel 4.13 Hasil Uji Autokorelasi

Nilai Uji Asumsi	
DL	1,5795
d	2.1497
DU	1,7549
P-value	0,696

Selanjutnya pada Tabel 4.13 didapatkan nilai Durbin Watson sebesar 2,1497 dengan *p-value* sebesar 0,696 > 0,05 terima H_0 sehingga dapat disimpulkan tidak terjadi autokorelasi pada data penelitian ini.

4.8 Uji Signifikansi Parameter

4.8.1 Uji Simultan

Uji Simultan dilakukan untuk mengetahui variabel independen bersama-sama dapat berpengaruh terhadap variabel dependen. Dengan hipotesis sebagai berikut:

Hipotesis :

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{10} = 0$ *vs*

$H_1 : \beta_1 \neq \beta_2 \neq \cdots \neq \beta_{10} \neq 0$ (minimal satu β_j tidak sama dengan 0)

Statistik Uji:

Berdasarkan hasil *R. 4.0.5* yang mengacu pada Lampiran 11 ndidapatkan nilai F-hitung dan *P-value* pada Tabel 4.14.

Tabel 4.14 Hasil Uji Simultan

Hasil Uji Simultan		Keputusan
F-hitung	F-tabel (0,05;10;58)	Tolak H_0
17,39	1,998	
P-value	Alfa	Tolak H_0
4.761e-14	0,05	

Nilai F-hitung dapat diperoleh menggunakan rumus sebagai berikut :

$$\text{F-hitung}=\frac{R^2/k}{(1-R^2)/(n-k-1)}=\frac{0,7499/10}{(1-7499)/58}=\frac{0,07499}{0,004312}=17,39$$

Berdasarkan hasil tabel 4.14 dan perhitungan manual mendapatkan hasil yang sama yaitu menolak H_0 dengan tingkat signifikansi 0,05 sehingga dapat disimpulkan bahwa variabel independen secara simultan berpengaruh terhadap curah hujan di DKI Jakarta.

4.8.2 Uji Parsial

Uji parsial dilakukan untuk menjukkan seberapa jauh pengaruh satu variabel independen secara parsial berpengaruh terhadap variabel dependen.

Dengan hipotesis sebagai berikut:

Hipotesis:

$H_0 : \beta_j = 0$ vs

$H_1: \beta_j \neq 0$ (minimal satu β_j tidak sama dengan 0)

Dimana j= 1,2,3,..k

Statistik Uji:

Berdasarkan hasil *R. 4.0.5* yang mengacu pada Lampiran 11 didapatkan nilai t-hitung pada Tabel 4.15.

Tabel 4.15 Hasil Uji Parsial

Parameter	T_{hitung}	*Ttabel* *(0,05;58)*	Keputusan
β_0 (Intersep)	3,574	2,0017	Tolak H_0
β_1 (X_1)	-1,381		Terima H_0
β_2 (X_2)	-3,417		Tolak H_0
β_3 $({X_2}^2)$	3,768		Tolak H_0
β_4 (X_3)	-3,511		Tolak H_0
β_5 (X_4)	-3,236		Tolak H_0
β_6 $({X_4}^2)$	3,195		Tolak H_0
β_7 $(X_1 \times X_2 \times X_3 \times X_4 \times \sqrt{X_1})$	2,789		Tolak H_0
β_8 $(X_1 \times {X_2}^2 \times X_3 \times X_4 \times \sqrt{X_1})$	-1,761		Terima H_0
β_9 $(X_1 \times X_2 \times X_3 \times {X_4}^2 \times \sqrt{X_1})$	-2,228		Tolak H_0
β_{10} $(X_1 \times {X_2}^2 \times X_3 \times {X_4}^2 \times \sqrt{X_1})$	1,471		Terima H_0

Berdasarkan Tabel 4.15 dapat dilihat bahwa variabel independen yang berpengaruh signifikan terhadap curah hujan di DKI Jakarta secara parsial adalah lama penyinaran matahari dan kecepatan angin rata-rata.

4.9 Kebaikan Model Regresi

Akurasi Model Regresi dapat dilihat dari nilai $R^2_{adjusted}$, *RSE, MSE*, dan nilai *RMSE*. Dapat dilihat kebaikan metode MKT dan *Robust Weighted Least Square* pada Tabel 4.16.

Tabel 4.16 Kebaikan Metode MKT dan Robust Weighted Least Square pada Regresi Non linier

Metode	$R^2_{adjusted}$	RSE	MSE	RMSE
MKT	36,35%	21,330	421,440	20,530
Robust Weighted Least Square	70,68%	0,795	0,532	0,729

Berdasarkan tabel 4.16 dapat diketahui metode *Robust Weighted Least Square* memiliki nilai $\boldsymbol{R^2_{adjusted}}$ lebih besar dibanding metode MKT. sedangkan nilai RSE, MSE, RMSE pada metode *Robust Weighted Least Square* lebih kecil maka tingkat akurasi data semakin tinggi dan model dikatakan lebih baik dibanding metode MKT.

4.10 Interpretasi Model

Berdasarkan subbab 4.7 didapatkan model *Robust Weighted Least Square* sebagai berikut:

$$\begin{aligned} y_{lts} = 92{,}73 - 0{,}3506X_1 - 2{,}028X_2 + 0{,}01364X_2^2 - 1{,}131X_3 \\ - 6{,}618X_4 + 1{,}495X_4^2 + 0{,}0002161X_1X_2X_3X_4\sqrt{X_1} \\ - 0{,}0000015535X_1X_2^2X_3X_4\sqrt{X_1} \\ - 0{,}00006108X_1X_2X_3{X_4}^2\sqrt{X_1} \\ - 0{,}0000004777X_1X_2^2X_3{X_4}^2\sqrt{X_1} \end{aligned}$$

Model tersebut memiliki $R^2_{adjusted}$ sebesar 70,68% yang artinya 70,68% keragaman curah hujan dipengaruhi oleh temperatur rata-rata, kelembapan rata-rata, lamanya penyinaran matahari, dan kecepatan angin rata-rata sedangkan 29,32% dijelaskan oleh variabel lain diluar analisis ini.

Setelah dilakukan pengujian kembali terhadap uji normalitas residual, heteroskedastisitas, dan non-autokorelasi diperoleh hasil sudah memenuhi asumsi terkecuali asumsi normalitas residual. Namun metode *robust weighted least square* tetap dapat mengatasi adanya *outlier* yang diindikasikan dengan mengecilnya nilai *RSE* dan *RMSE*.

Berdasarkan subbab 4.8.2, variabel independen yang signifikan terhadap curah hujan adalah lama penyinaran matahari dan kecepatan angin rata-rata maka interpretasi dari persamaan regresi *robust weighted least square* metode LTS sebagai berikut:

1. Intersep 92,73 artinya apabila variabel independen konstan maka rata-rata curah hujan di DKI Jakarta sebesar 92,73 mm.
2. Koefisien regresi kelembapan rata-rata bernilai negatif yang artinya pada saat kelembapan rata-rata meningkat 1 persen maka akan menurunkan curah hujan sebesar 2,028 mm.
3. Koefisien regresi kelembapan rata-rata orde 2 bernilai positif yang artinya pada saat kelembapan rata-rata orde 2 meningkat satu persen maka curah hujan akan meningkat secara kuadratik membentuk garis lengkung.
4. Koefisien regresi lama penyinaran matahari bernilai negatif yang artinya pada saat lama penyinaran matahari meningkat 1 jam maka akan menurunkan curah hujan sebesar 1,131 mm.
5. Koefisien regresi kecepatan angin rata-rata bernilai negatif yang artinya pada saat kecepatan angin rata-rata meningkat 1 m/s maka akan menurunkan curah hujan sebesar 6,618 mm.
6. Koefisien regresi kecepatan angin rata-rata orde 2 bernilai positif yang artinya pada saat kecepatan angin rata-rata orde 2 meningkat 1 m/s maka curah hujan akan meningkat secara kuadratik membentuk garis lengkung.

Indonesia sebagai Negara tropis memiliki kelembapan udara yang relatif tinggi, Hal ini dapat terlihat dari persentase kelembapan udara maksimum yang mencapai 93% dari bulan November 2019 hingga awal Februari 2020 di DKI Jakarta yang diukur dari Stasiun Meteorologi Maritim Tanjung Priok. Dalam kurun waktu yang sama, rata-rata persentase kelembapan udara adalah sebesar 78,97% dengan kelembapan minimumnya adalah 64% yang terjadi pada tanggal 9 November 2019. Berdasarkan hasil penelitian kelembapan rata-rata orde 1 dan orde 2 berpengaruh signifikan terhadap curah hujan. Dimana pada orde 1 kelembapan rata-rata berpengaruh signifikan secara negatif terhadap curah hujan sedangkan pada orde 2 kelembapan udara berpengaruh secara positif dengan peningkatannya secara kuadratik.

Lama penyinaran matahari merupakan salah satu unsur klimatologi yang perlu dipantau karena dapat mengindikasikan terjadinya perubahan iklim. Dari bulan November 2019 hingga awal Februari 2020, penyinaran matahari paling lama 10 jam dalam sehari. Dalam kurun waktu yang sama lama penyinaran matahari paling sebentar selama 0,1 jam yang kemungkinan cuaca pada hari tersebut mendung. Lama penyinaran matahari yang memiliki durasi pendek dapat menjadi penduga mengenai kejadian hujan yang terjadi di bulan tersebut. Berdasarkan hasil penelitian lama penyinaran matahari memiliki pengaruh yang signifikan terhadap curah hujan. Lama penyinaran matahari memiliki pengaruh yang negatif terhadap curah hujan yang menandakan semakin lama penyinaran matahari maka semakin kecil peluang kemungkinan terjadinya hujan.

Kecepatan angin rata-rata memiliki pengaruh yang signifikan terhadap curah hujan. Dari bulan November 2019 hingga awal Februari 2020, kecepatan angin rata-rata paling rendah adalah sebesar 1 m/s yang sering terjadi pada hari-hari tertentu dalam kurun waktu tersebut. Berdasarkan hasil penelitian kecepatan angin rata-rata orde 1 memiliki pengaruh signifikan secara negatif terhadap curah hujan sedangkan kecepatan angin rata-rata orde dua memiliki pengaruh signifikan secara positif terhadap curah hujan yang peningkatannya secara kuadratik.

BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan maka dapat diperoleh kesimpulan sebagai berikut:

1. Heteroskedastisitas pada data curah hujan di DKI Jakarta dapat diatasi dengan menggunakan *Weighted Least Square.*
2. Pencilan pada data curah hujan di DKI Jakarta dapat diatasi menggunakan regresi *Robust* metode *Least Trimmed Square.* Pada penelitian ini, asumsi normalitas residual masih belum terpenuhi namun heteroskedastisitas dan pencilan dapat diatasi menggunakan *Robust Weighted Least Square* dengan nilai $R^2_{adjusted}$ sebesar 70,68% dan nilai *RMSE* sebesar 0,729.

5.2 Saran

Berdasarkan penelitian yang telah dilakukan, dapat diberikan beberapa saran yaitu:

1. Bagi penelitian selanjutnya disarankan untuk menggunakan studi kasus selain data curah hujan selain di DKI Jakarta untuk mengekplorasi lebih luas penerapan regresi non-linier.
2. Pada penelitian selanjutnya diharapkan dapat mengatasi normalitas residual pada analisis regresi nonlinier menggunakan *Robust Weighted Least Square.*

DAFTAR PUSTAKA

Chatterjee, S., & Simonoff, J. S. (2013). *Handbook of Regression Analysis*. Canada: John Wiley & Sons, Inc.

Chen, C. (2002). *Robust Regression and Outlier Detection with ROBUSTREG Procedure.* SAS Institute Inc.

Drapper, N., & Smith, H. (1992). *Analisis Regresi Terapan. Alih Bahasa oleh Sumantri, B. PT Gramedia Pustaka Utama.* Jakarta: PT. Gramedia Pustaka Utama.

Ghozali, I. (2013). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 21 Update PLS Regresi*. Semarang: Badan Penerbit Universitas Diponegoro.

Gujarati, D. N. (2003). *Basic Econometrics, Fourth Edition.* New York: Mc. Graw Hill.

Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics. Fifth Edition.* New York: McGraw Hill.

Kutner, M. H., Neter, J., & Li, W. (2004). *Applied Linier Regression Models. Fifth Edition.* . New York: McGraw-Hill Companies, Inc.

Montgomery, D. C., & Peck, E. A. (1992). *Introduction To Linear Regressio Analysis. Second Edition.* New York: John Wiley and Sons, Inc.

Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool, Second Edition.* New York: Springer-Verlag.

Soemartini. (2007). *Outlier (outlier).* Bandung: Universitas Padjadjaran Wordpress.

Steel, R. G., & Torrie, J. (1980). *Principles and Procedure of Statistics, Second Edition.* McGraw Hill Book Company.

Tanudidjaja. (1993). *Ilmu Pengetahuan Bumi dan Antariksa*. Jakarta: Penerbit Departemen Pendidikan dan Kebudayaan.

Tjasyono. (1999). *Klimatologi Umum.* Bandung: ITB.

LAMPIRAN

Lampiran 1. Data Curah Hujan DKI Jakarta Periode November – awal Februari

Tanggal	Tavg	RH_avg	ss	ff_avg	RR
01-11-2019	28,7	74	7,7	2	0
02-11-2019	29,2	70	6,5	2	29,1
03-11-2019	29,2	71	9,3	1	
04-11-2019	29,3	70	9	2	
05-11-2019	29,5	73	4,6	2	8888
06-11-2019	29,2	71	9,3	2	27,5
07-11-2019	28,6	72	8,6	1	
08-11-2019	29	74	9,6	2	
09-11-2019	29,6	64	9,4	2	
10-11-2019	29,5	72	9,6	2	0
11-11-2019	30	71	9,3	2	0
12-11-2019	30,1	72	9,7	1	
13-11-2019	30,6	70	5,6	2	
14-11-2019	28,5	77	2,9	3	
15-11-2019	30,3	72	0,6	2	2,5
16-11-2019	29,9	70	9,6	2	0
17-11-2019	29,8	70	9,4	2	0
18-11-2019	29,7	72	9,6	2	
19-11-2019	29,6	74	7,1	2	
20-11-2019	30,1	70	9,5	2	0
21-11-2019	29,2	76	8,8	2	
22-11-2019	29,9	73	6,2	1	0,3
23-11-2019	29,9	73	6,9	2	
24-11-2019	29,7	72	3,7	2	
25-11-2019	29,8	76	9,4	2	6,5
26-11-2019	30	74	6,9	2	0
27-11-2019	30,1	74	8,3	2	
28-11-2019	30	74	6,7	2	

Tanggal	Tavg	RH_avg	ss	ff_avg	RR
29-11-2019	30,1	74	9,5	2	
30-11-2019	28	82	7,1	2	2,1
01-12-2019	30,1	75	6,6	2	12
02-12-2019	30,2	73	9,4	1	3,5
03-12-2019	30,3	74	8	2	
04-12-2019	29,8	76	9,5	3	
05-12-2019	29,4	78	9,6	2	10
06-12-2019	28,9	80	7,4	3	8888
07-12-2019	27,7	84	0,6	2	4,3
08-12-2019	28,3	83	0,2	2	11
09-12-2019	29,3	74	9,3	2	
10-12-2019	28,7	78	7,7	2	0
11-12-2019	28,5	74	2	2	0
12-12-2019	27,6	82		2	
13-12-2019	28,5	78		2	3,4
14-12-2019	28,9	75	2,8	1	
15-12-2019	29,3	76	8,6	2	
16-12-2019	29	80	6,1	2	0
17-12-2019	27,3	84	0,9	1	3,3
18-12-2019	28,7	78	3,7	2	
19-12-2019	28,6	80	9	1	
20-12-2019	29,1	77	5,5	1	8888
21-12-2019	29,6	76	8,1	2	0
22-12-2019	27,6	85	8,5	2	54
23-12-2019	28,4	81	5,6	2	5,5
24-12-2019	29	77	6,6	2	7,6
25-12-2019	26,6	93	8,5	1	102,5
26-12-2019	28,2	82		2	44
27-12-2019	27,6	86	6,8	1	2,5
28-12-2019	27,5	83	4,8	1	50,2
29-12-2019	27,2	86	5,3	1	13,2
30-12-2019	29	79		2	1,2

Tanggal	Tavg	RH_avg	ss	ff_avg	RR
31-12-2019	27,6	84	9,6	4	0
01-01-2020	27,4	88	3,1	3	146,1
02-01-2020	27,9	83		2	13,5
03-01-2020	28,6	79	7,3	3	0
04-01-2020	28,2	85	0,8	3	6,4
05-01-2020	27,8	86	0,8	3	1,1
06-01-2020	27,2	87		3	54
07-01-2020	27,3	88	3,2	3	16,9
08-01-2020	28,1	80	4,7	3	0,2
09-01-2020	28,4	82	0,1	4	69,4
10-01-2020	25,9	88		2	20,1
11-01-2020	26,6	86		3	21
12-01-2020	28,3	79	2,2	3	4,5
13-01-2020	27,4	87	10	3	0,4
14-01-2020	28,9	78	2,2	3	0,3
15-01-2020	28,2	82	8,5	1	35
16-01-2020	28,5	82	5,4	2	
17-01-2020	28,9	79	4	3	1,1
18-01-2020	27,4	84	9,1	2	23,5
19-01-2020	28,5	81	2,8	2	2,2
20-01-2020	29	80	9,4	2	
21-01-2020	28,7	82	4,9	2	
22-01-2020	28,8	81	6,3	2	8888
23-01-2020	28,1	86	7	2	10,6
24-01-2020	27	91	5,5	2	19,9
25-01-2020	27,5	87		2	112,3
26-01-2020	28,3	84	0,6	2	2
27-01-2020	28,9	80	8,1	2	1,5
28-01-2020	28,8	82		2	3,7
29-01-2020	29,6	81	8,4	2	
30-01-2020	27,3	82	7,8	1	41,5
31-01-2020	28,5	82	3,3	2	0

Tanggal	Tavg	RH_avg	ss	ff_avg	RR
01-02-2020	26,1	93	0,4	1	42,9
02-02-2020	26,7	90	0,2	1	86,1
03-02-2020	27	89	2,5	1	39,3

Keterangan:
8888 : data tidak terukur
Tavg : Temperatur rata-rata (ºC)
RH_avg : Kelembapan rata-rata (%)
ss : lamanya penyinaran matahari (jam)
ff_avg : kecepatan angin rata-rata (m/s)
RR : curah hujan (mm)

Lampiran 2. *Output* Analisis Deskriptif

```
> #read data
> dat<-read_excel("E:/ratih1.xlsx",sheet=6)
> summary(dat)
      Tavg           RH_avg            ss              ff_avg     
 Min.   :25.90   Min.   :64.00   Min.   : 0.100   Min.   :1.000  
 1st Qu.:27.95   1st Qu.:74.00   1st Qu.: 3.700   1st Qu.:2.000  
 Median :28.70   Median :79.00   Median : 6.900   Median :2.000  
 Mean   :28.67   Mean   :78.97   Mean   : 6.232   Mean   :2.011  
 3rd Qu.:29.55   3rd Qu.:83.00   3rd Qu.: 9.000   3rd Qu.:2.000  
 Max.   :30.60   Max.   :93.00   Max.   :10.000   Max.   :4.000  
                                 NA's   :10                      
       RR        
 Min.   :  0.00  
 1st Qu.:  0.30  
 Median :  4.30  
 Mean   : 18.60  
 3rd Qu.: 22.25  
 Max.   :146.10  
 NA's   :32      
> head(dat)
# A tibble: 6 x 5
   Tavg RH_avg    ss ff_avg    RR
  <dbl>  <dbl> <dbl>  <dbl> <dbl>
1  28.7     74   7.7      2   0  
2  29.2     70   6.5      2  29.1
3  29.2     71   9.3      1  NA  
4  29.3     70   9        2  NA  
5  29.5     73   4.6      2  NA  
6  29.2     71   9.3      2  27.5
```

Lampiran 3. *Output* Regresi Polinomial *Software R*

```
> l1=RRT(X1, Y)
> l1
          R-squared 1 R-squared 2     F p-value
Linear          0.257       0.306 0.916   0.404
Quadratic       0.000       0.000 0.640   0.591
Cubic           0.000       0.000 1.603   0.180
> l2=RRT(X2, Y)
> l2
          R-squared 1 R-squared 2      F p-value
Linear          0.341       0.346 10.709   0.000
Quadratic       0.000       0.000  0.055   0.983
Cubic           0.000       0.000  0.180   0.948
> l3=RRT(X3, Y)
> l3
          R-squared 1 R-squared 2     F p-value
Linear          0.037       0.048 0.086   0.917
Quadratic       0.000       0.000 0.393   0.758
Cubic           0.000       0.000 0.267   0.898
> l4=RRT(X4, Y)
> l4
          R-squared 1 R-squared 2     F p-value
Linear          0.078       0.078 3.554   0.033
Quadratic       0.000       0.000 0.089   0.966
Cubic           0.000       0.000 0.000   1.000
```

Lampiran 4. Deteksi Multikolinieritas

```
> #deteksi Multikolinieritas
> cor(data1[1:4])
             Tavg       RH_avg         ss      ff_avg
Tavg    1.0000000 -0.87967640  0.4963874 -0.01320060
RH_avg -0.8796764  1.00000000 -0.4720335  0.07158367
ss      0.4963874 -0.47203348  1.0000000 -0.17196737
ff_avg -0.0132006  0.07158367 -0.1719674  1.00000000
> library(car)
Loading required package: carData
> reg = lm(Y~X1+X2+X3+X4,data1)
> vif(reg)
      X1       X2       X3       X4 
4.684970 4.511001 1.382509 1.051108 
```

Lampiran 5. Uji Homoskedastisitas

```
> regresi = lm(Y~X1+X2+X3+X4+X2q+X4q,data2)
>
> summary(regresi)

Call:
lm(formula = Y ~ X1 + X2 + X3 + X4 + X2q + X4q, data = data2)

Residuals:
    Min      1Q  Median      3Q     Max
-37.804  -9.554  -4.032   4.124 107.321

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1408.24516  339.25910   4.151 7.63e-05 ***
X1            -4.48931    4.83025  -0.929 0.355216
X2           -32.82923    8.64752  -3.796 0.000269 ***
X3             0.58492    0.82098   0.712 0.478056
X4           -18.26474   15.24291  -1.198 0.234039
X2q            0.21395    0.05559   3.849 0.000225 ***
X4q            4.29932    3.44420   1.248 0.215240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 21.33 on 88 degrees of freedom
Multiple R-squared:  0.3635,    Adjusted R-squared:  0.3201
F-statistic: 8.376 on 6 and 88 DF,  p-value: 3.442e-07

> res <- resid(regresi)
>
> #Uji Homokedastisitas
> bptest(regresi)

        studentized Breusch-Pagan test

data:  regresi
BP = 17.515, df = 6, p-value = 0.007565

> #ASUMSI HOMOKEDASTISITAS TIDAK TERPENUHI
```

Lampiran 6. Uji Normalitas

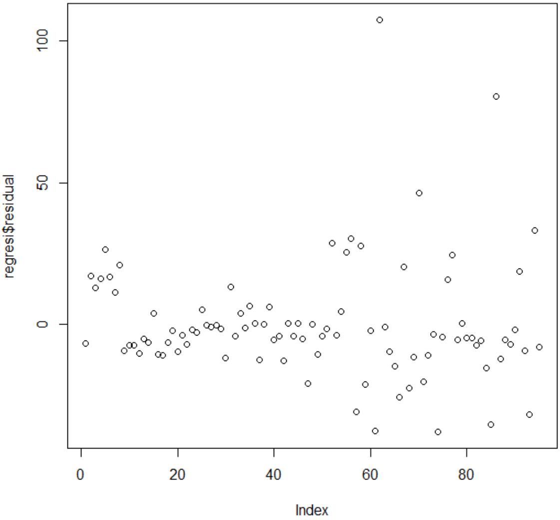
```
> #Uji Normalitas
> residual=resid(regresi)
> lillie.test(residual)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  residual
D = 0.21956, p-value = 2.934e-12

> #ASUMSI NORMALITAS TIDAK TERPENUHI
```

Lampiran 7. Uji Non-autokorelasi

```
> #Uji Non Autokorelasi
> library(lmtest)
> dwtest(regresi)

        Durbin-Watson test

data:  regresi
DW = 2.4417, p-value = 0.979
alternative hypothesis: true autocorrelation is greater than 0

> #ASUMSI NON AUTOKORELASI TERPENUHI
```

Lampiran 8. *Output* Deteksi Pencilan

Pencilan : 10,53%




```
> #DETEKSI OUTLIER (Plot Residual)
> plot(regresi$residual)
>
> #DETEKSI OUTLIER (DFFITS)
> dffits(regresi)
           1            2            3            4            5
-0.0774602865  0.2362279171  0.2033336478  0.2170760434  0.2343881349
           6            7            8            9           10
 0.2042266716  0.2057615833  0.2194861241 -0.3867441151 -0.0697033799
          11           12           13           14           15
-0.0721993318 -0.1620054452 -0.0815623066 -0.0820115430  0.0775191924
          16           17           18           19           20
-0.1220462553 -0.1257669946 -0.0581707899 -0.0161975747 -0.1123679537
          21           22           23           24           25
-0.0289032694 -0.1095070980 -0.0161271726 -0.0315178026  0.0495674772
          26           27           28           29           30
-0.0042191243 -0.0079889714 -0.0032153922 -0.0152786917 -0.1159380264
          31           32           33           34           35
 0.1388361821 -0.0717277430  0.0416483730 -0.0154838243  0.0644302250
          36           37           38           39           40
 0.0037872113 -0.1439418906 -0.0004092769  0.0496456100 -0.0469737909
          41           42           43           44           45
-0.0663281621 -0.1579943408  0.0032772748 -0.0669986734  0.0019490570
          46           47           48           49           50
-0.0418160064 -0.3411482278  0.0008328339 -0.1562080961 -0.0590647862
          51           52           53           54           55
-0.0116569436  0.3567294609 -0.0285134095  0.0303900543  0.8134949711
          56           57           58           59           60
 0.3152129395 -0.4097544927  0.3945001172 -0.2751358321 -0.0177339156
          61           62           63           64           65
-1.768568363  1.7480911841 -0.0091378618 -0.1066402041 -0.1998767478
          66           67           68           69           70
-0.3300586964  0.2632156293 -0.3004357501 -0.1312963904  2.0782990608
          71           72           73           74           75
-0.4350847569 -0.1807370217 -0.0439992645 -0.7147640997 -0.0511844627
          76           77           78           79           80
 0.2237729576  0.1884985397 -0.0546908272  0.0048283008 -0.0444625457
          81           82           83           84           85
-0.0504881952 -0.0620756641 -0.0452396260 -0.1731072678 -0.7393179199
          86           87           88           89           90
 0.9382369780 -0.1509993835 -0.0460572789 -0.0717963335 -0.0258423112
          91           92           93           94           95
 0.3456694636 -0.0823700506 -0.9111023862  0.6334564370 -0.1252955408
>
```

Lampiran 9. Output WLS

```
> #WLS
> X2=cbind(X2,X2^2)
> X4=cbind(X4,X4^2)
> WLS1 = lm (Y/X1~(X1+X2+X3+X4)/X1,data2)
> WLS2 = lm (Y/X2~(X1+X2+X3+X4)/X2,data2)
Error in Y/X2 : non-conformable arrays
> WLS3 = lm (Y/X3~(X1+X2+X3+X4)/X3,data2)
> WLS4 = lm (Y/X4~(X1+X2+X3+X4)/X4,data2)
Error in Y/X4 : non-conformable arrays
> WLS5 = lm (Y/sqrt(X1)~(X1+X2+X3+X4)/sqrt(X1),data2)
> WLS6 = lm (Y/sqrt(X2)~(X1+X2+X3+X4)/sqrt(X2),data2)
Error in Y/sqrt(X2) : non-conformable arrays
> WLS7 = lm (Y/sqrt(X3)~(X1+X2+X3+X4)/sqrt(X3),data2)
> WLS8 = lm (Y/sqrt(X4)~(X1+X2+X3+X4)/sqrt(X4),data2)
Error in Y/sqrt(X4) : non-conformable arrays
```

Lampiran 9. Lanjutan (*Output WLS*)

```
> bptest(WLS1)

        studentized Breusch-Pagan test

data:  WLS1
BP = 18.243, df = 10, p-value = 0.05099

> bptest(WLS2)
Error in bptest(WLS2) : object 'WLS2' not found
> bptest(WLS3)

        studentized Breusch-Pagan test

data:  WLS3
BP = 55.148, df = 10, p-value = 2.963e-08

> bptest(WLS4)
Error in bptest(WLS4) : object 'WLS4' not found
> bptest(WLS5)

        studentized Breusch-Pagan test

data:  WLS5
BP = 17.911, df = 10, p-value = 0.05649

> bptest(WLS6)
Error in bptest(WLS6) : object 'WLS6' not found
> bptest(WLS7)

        studentized Breusch-Pagan test

data:  WLS7
BP = 35.337, df = 10, p-value = 0.0001094

> bptest(WLS8)
Error in bptest(WLS8) : object 'WLS8' not found
```

Lampiran 9. Lanjutan (*Output WLS*)

```
> summary(WLS1)

Call:
lm(formula = Y/X1 ~ (X1 + X2 + X3 + X4)/X1, data = data2)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.4826 -0.3114 -0.1183  0.2018  3.8736 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)    6.229e+01  1.522e+01   4.094  9.7e-05 ***
X1            -2.557e-01  1.779e-01  -1.437 0.154389    
X21           -1.362e+00  3.477e-01  -3.918 0.000181 ***
X22            8.638e-03  2.130e-03   4.056 0.000111 ***
X3            -7.506e-02  1.717e-01  -0.437 0.663139    
X41           -1.645e+00  1.029e+00  -1.598 0.113854    
X42            4.619e-01  2.212e-01   2.088 0.039782 *  
X1:X21:X3:X41 -8.369e-05  2.491e-04  -0.336 0.737732    
X1:X22:X3:X41  2.101e-06  2.855e-06   0.736 0.463904    
X1:X21:X3:X42  2.099e-05  9.251e-05   0.227 0.821098    
X1:X22:X3:X42 -5.926e-07  1.079e-06  -0.549 0.584503    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7422 on 84 degrees of freedom
Multiple R-squared:  0.4506,	Adjusted R-squared:  0.3852 
F-statistic:  6.89 on 10 and 84 DF,  p-value: 8.975e-08
```

Lampiran 9. Lanjutan (*Output WLS*)

```
> summary(WLS5)

Call:
lm(formula = Y/sqrt(X1) ~ (X1 + X2 + X3 + X4)/sqrt(X1), data = data2)

Residuals:
    Min      1Q  Median      3Q     Max 
-7.7087 -1.6884 -0.6152  1.0829 20.3065 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)              3.302e+02  8.259e+01   3.998 0.000137 ***
X1                      -1.410e+00  9.643e-01  -1.462 0.147448    
X21                     -7.170e+00  1.859e+00  -3.858 0.000224 ***
X22                      4.533e-02  1.134e-02   3.996 0.000138 ***
X3                      -3.402e-01  8.940e-01  -0.381 0.704475    
X41                     -8.387e+00  5.379e+00  -1.559 0.122705    
X42                      2.386e+00  1.159e+00   2.060 0.042519 *  
X1:X21:X3:X41:sqrt(X1) -1.094e-04  2.400e-04  -0.456 0.649804    
X1:X22:X3:X41:sqrt(X1)  2.364e-06  2.801e-06   0.844 0.401064    
X1:X21:X3:X42:sqrt(X1)  2.836e-05  9.006e-05   0.315 0.753606    
X1:X22:X3:X42:sqrt(X1) -6.726e-07  1.057e-06  -0.636 0.526202    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.902 on 84 degrees of freedom
Multiple R-squared:  0.444,	Adjusted R-squared:  0.3779 
F-statistic: 6.709 on 10 and 84 DF,  p-value: 1.399e-07
```

Lampiran 10. Output Asumsi Normalitas pada model WLS

```
> #Uji Normalitas
> residual1=resid(WLS1)
> residual2=resid(WLS5)
> lillie.test(residual1)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  residual1
D = 0.14166, p-value = 6.993e-05

> lillie.test(residual2)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  residual2
D = 0.14199, p-value = 6.627e-05
```

Lampiran 11. *Output Robust LTS*

```
> regresilts = ltsReg(Y/(X1)~(X1+X2+X3+X4)/(X1),data2)
Warning message:
In covMcd(X, alpha = alpha, use.correction = use.correction) :
  The 53-th order statistic of the absolute deviation of variable 5 is
zero.
There are 60 observations (in the entire dataset of 95 obs.) lying on
the hyperplane with equation a_1*(x_i1 - m_1) + ... + a_p*(x_ip - m_p)
= 0 with (m_1, ..., m_p) the mean of these observations and
coefficients a_i from the vector a <- c(0, 0, 0, 0, 1, 0, 0, 0, 0, 0)
> summary(regresilts)

Call:
ltsReg.formula(formula = Y/(X1) ~ (X1 + X2 + X3 + X4)/(X1), data = data2)

Residuals (from reweighted LS):
     Min       1Q   Median       3Q      Max
-0.13334 -0.04295  0.00000  0.01673  0.36811

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
Intercept     -3.857e+00  5.350e+00  -0.721    0.474
X1             5.670e-02  4.097e-02   1.384    0.172
X21            4.373e-02  1.293e-01   0.338    0.736
X22           -1.668e-04  8.121e-04  -0.205    0.838
X3            -7.040e-03  4.900e-02  -0.144    0.886
X41           -4.873e-02  3.256e-01  -0.150    0.882
X42            3.239e-03  7.520e-02   0.043    0.966
X1:X21:X3:X41 -2.570e-05  7.150e-05  -0.359    0.721
X1:X22:X3:X41  3.319e-07  7.674e-07   0.432    0.667
X1:X21:X3:X42  1.331e-05  2.427e-05   0.549    0.585
X1:X22:X3:X42 -1.646e-07  2.737e-07  -0.602    0.550

Residual standard error: 0.1272 on 57 degrees of freedom
Multiple R-Squared: 0.1614,    Adjusted R-squared: 0.01433
F-statistic: 1.097 on 10 and 57 DF,  p-value: 0.3798
```

Lampiran 11. Lanjutan (*Output Robust LTS*)

```
> regresiltsl= ltsReg(Y/sqrt(X1)~(X1+X2+X3+X4)/sqrt(X1),data2)
Warning message:
In covMcd(X, alpha = alpha, use.correction = use.correction) :
  The 53-th order statistic of the absolute deviation of variable 5 is
zero.
There are 60 observations (in the entire dataset of 95 obs.) lying on
the hyperplane with equation a_1*(x_i1 - m_1) + ... + a_p*(x_ip - m_p)
= 0 with (m_1, ..., m_p) the mean of these observations and
coefficients a_i from the vector a <- c(0, 0, 0, 0, 1, 0, 0, 0, 0, 0)
> summary(regresiltsl)

Call:
ltsReg.formula(formula = Y/sqrt(X1) ~ (X1 + X2 + X3 + X4)/sqrt(X1),
    data = data2)

Residuals (from reweighted LS):
    Min      1Q  Median      3Q     Max
-1.5041 -0.2174  0.0000  0.2003  2.4941

Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
Intercept               9.273e+01  2.595e+01   3.574 0.000717 ***
X1                     -3.506e-01  2.538e-01  -1.381 0.172461
X21                    -2.028e+00  5.935e-01  -3.417 0.001162 **
X22                     1.364e-02  3.619e-03   3.768 0.000387 ***
X3                     -1.131e+00  3.223e-01  -3.511 0.000872 ***
X41                    -6.618e+00  2.045e+00  -3.236 0.002005 **
X42                     1.495e+00  4.681e-01   3.195 0.002265 **
X1:X21:X3:X41:sqrt(X1)  2.161e-04  7.748e-05   2.789 0.007133 **
X1:X22:X3:X41:sqrt(X1) -1.553e-06  8.816e-07  -1.761 0.083502 .
X1:X21:X3:X42:sqrt(X1) -6.108e-05  2.741e-05  -2.228 0.029771 *
X1:X22:X3:X42:sqrt(X1)  4.777e-07  3.248e-07   1.471 0.146732
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7954 on 58 degrees of freedom
Multiple R-Squared: 0.7499,     Adjusted R-squared: 0.7068
F-statistic: 17.39 on 10 and 58 DF,  p-value: 4.761e-14
```

Lampiran 12. *Output* Pengujian Asumsi

```
> residual4=resid(regresiltsl)
>
> lillie.test(residual4)

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  residual4
D = 0.27608, p-value < 2.2e-16

>
>
> bptest(regresiltsl)

        studentized Breusch-Pagan test

data:  regresiltsl
BP = 17.911, df = 10, p-value = 0.05649

>
>
> dwtest(regresiltsl)

        Durbin-Watson test

data:  regresiltsl
DW = 2.1497, p-value = 0.696
alternative hypothesis: true autocorrelation is greater than 0
```

Lampiran 13. *Syntax software R*

```
clear(workspace)

#packages
library(readxl)
library(lmtest)
library(ggplot2)
library(nortest)
library(tseries)
library(MASS)
library(zoo)
library(forecast)


#read data
dat<-read_excel("E:/ratih1.xlsx",sheet=6)
summary(dat)
head(dat)
data=na.locf(dat)
data1=data.frame(data)
X1=cbind(data1$Tavg)
X2=cbind(data1$RH_avg)
X3=cbind(data1$ss)
X4=cbind(data1$ff_avg)
Y=cbind(data1$RR)
```

```
head(data1)

nrow(data1)

summary(data1)


#deteksi Multikolinieritas

cor(data1[1:4])

library(car)

reg = lm(Y~X1+X2+X3+X4,data1)

vif(reg)

#NON MULTIKOLINIERITAS TERPENUHI


#Uji Linier (NON LINIER)

Y = as.matrix(data1[,5])

X = as.matrix(data1[,1:4])

RRT = function(X,Y){

library(lmtest)

RESET = matrix(0,3,4)

colnames(RESET) = c("R-squared 1","R-squared 
2","F","p-value")

rownames(RESET) = c("Linear","Quadratic","Cubic")

#linearmodel

mod.a = lm(Y~X)

Rsq.a = 1-((sum((Y-fitted(mod.a))^2))/(sum((Y-
mean(Y))^2)))

XY = cbind(X,fitted(mod.a)^2)
```

```
mod.b = lm(Y~XY)

Rsq.b = 1 -((sum((Y-fitted(mod.b))^2))/(sum((Y-
mean(Y))^2)))

db1 = 2

db2 = length(X)-
length(as.matrix(summary(mod.b)$coeff[,1]))

RESET[1,1] = Rsq.a

RESET[1,2] = Rsq.b

RESET[1,3] = ((Rsq.b - Rsq.a)/db1) / ((1 -
Rsq.b)/db2)

RESET[1,4] = pf(RESET[1,3],db1,db2,lower.tail = F)

# quadratic model

X2 = cbind(X, X^2)

mod.a = lm (Y~X2)

Rsq.a = 1 - ((sum((Y - fitted(mod.a))^2)) / (sum((Y
- mean(Y))^2)))

X2Y = cbind(X2, fitted(mod.a)^2)

mod.b = lm (Y~X2Y)

Rsq.b = 1 - ((sum((Y - fitted(mod.b))^2)) / (sum((Y
- mean(Y))^2)))

db1  = 3

db2  = length(X) -
length(as.matrix(summary(mod.b)$coeff[,1]))

RESET[1,1] = Rsq.a

RESET[1,2] = Rsq.b

RESET[2,3] = ((Rsq.b - Rsq.a)/db1) / ((1 -
Rsq.b)/db2)
```

```
RESET[2,4] = pf(RESET[2,3],db1,db2,lower.tail = F)

# cubic model

X3 = cbind(X, X^2, X^3)

mod.a = lm (Y~X3)

Rsq.a = 1 - ((sum((Y - fitted(mod.a))^2)) / (sum((Y
- mean(Y))^2)))

X3Y = cbind(X3, fitted(mod.a)^2)

mod.b = lm (Y~X3Y)

Rsq.b = 1 - ((sum((Y - fitted(mod.b))^2)) / (sum((Y
- mean(Y))^2)))

db1   = 4

db2   = length(X) -
length(as.matrix(summary(mod.b)$coeff[,1]))

RESET[1,1] = Rsq.a

RESET[1,2] = Rsq.b

RESET[3,3] = ((Rsq.b - Rsq.a)/db1) / ((1 -
Rsq.b)/db2)

RESET[3,4] = pf(RESET[3,3],db1,db2,lower.tail = F)

RRT = round(RESET,3)

return(RRT)}

l1=RRT(X1, Y)

l1

l2=RRT(X2, Y)

l2

l3=RRT(X3, Y)

l3
```

```
l4=RRT(X4, Y)

l4

resettest(Y~X1,type="regressor")

resettest(Y~X2,type="regressor")

resettest(Y~X3,type="regressor")

resettest(Y~X4,type="regressor")



data2=data.frame(data)

X1=cbind(data2$Tavg)

X2=cbind(data2$RH_avg)

X3=cbind(data2$ss)

X4=cbind(data2$ff_avg)

Y=cbind(data2$RR)

X2q=X2^2

X4q=X4^2

regresi = lm(Y~X1+X2+X3+X4+X2q+X4q,data2)


summary(regresi)

res <- resid(regresi)


#Uji Homokedastisitas

bptest(regresi)

#ASUMSI HOMOKEDASTISITAS TIDAK TERPENUHI
```

```
#Uji Normalitas

residual=resid(regresi)

lillie.test(residual)

#ASUMSI NORMALITAS TIDAK TERPENUHI


#DETEKSI OUTLIER (Plot Residual)

plot(regresi$residual)


#DETEKSI OUTLIER (DFFITS)

dffits(regresi)


#Uji Non Autokorelasi

library(lmtest)

dwtest(regresi)

#ASUMSI NON AUTOKORELASI TERPENUHI



#WLS

X2=cbind(X2,X2^2)

X4=cbind(X4,X4^2)

WLS1 = lm (Y/X1~(X1+X2+X3+X4)/X1,data2)

WLS2 = lm (Y/X2~(X1+X2+X3+X4)/X2,data2)

WLS3 = lm (Y/X3~(X1+X2+X3+X4)/X3,data2)

WLS4 = lm (Y/X4~(X1+X2+X3+X4)/X4,data2)
```

```
WLS5 = lm (Y/sqrt(X1)~(X1+X2+X3+X4)/sqrt(X1),data2)
WLS6 = lm (Y/sqrt(X2)~(X1+X2+X3+X4)/sqrt(X2),data2)
WLS7 = lm (Y/sqrt(X3)~(X1+X2+X3+X4)/sqrt(X3),data2)
WLS8 = lm (Y/sqrt(X4)~(X1+X2+X3+X4)/sqrt(X4),data2)
bptest(WLS1)
bptest(WLS2)
bptest(WLS3)
bptest(WLS4)
bptest(WLS5)
bptest(WLS6)
bptest(WLS7)
bptest(WLS8)
summary(WLS1)
summary(WLS5)


#Uji Normalitas
residual1=resid(WLS1)
residual2=resid(WLS5)
lillie.test(residual1)
lillie.test(residual2)
#ASUMSI NORMALITAS TIDAK TERPENUHI
```

```
library(robustbase)
X1=cbind(data2$Tavg)
X2=cbind(data2$RH_avg)
X3=cbind(data2$ss)
X4=cbind(data2$ff_avg)
Y=cbind(data2$RR)
X2=cbind(X2,X2^2)
X4=cbind(X4,X4^2)
regresilts =
ltsReg(Y/(X1)~(X1+X2+X3+X4)/(X1),data2)
summary(regresilts)
regresilts1=
ltsReg(Y/sqrt(X1)~(X1+X2+X3+X4)/sqrt(X1),data2)
summary(regresilts1)


residual4=resid(regresilts1)


lillie.test(residual4)



bptest(regresilts1)


dwtest(regresilts1)
```